

EXHIBIT 77

**TO THE DECLARATION OF THARAN GREGORY LANIER
IN SUPPORT OF DEFENDANTS'
MOTION FOR SUMMARY JUDGMENT**

Ovum Decision Matrix: Selecting an Analytic Database, 2013–14

Reference Code: IT014-002786

Publication Date: 14 Jan 2014

Author: Surya Mukherjee, Madan Sheina

SUMMARY

Catalyst

Over the last two to three years, the emergence of purpose-built and workload-optimized analytic databases has shaken up the previously commoditized database market. Relational databases have started to make headlines again thanks to the advent of Big Data and alternative "NoSQL" database infrastructures such as Hadoop. The primary use cases behind this renewed interest have arisen because the goalposts for business analytics are constantly shifting, which increasingly renders traditional approaches inadequate. Most notably, skyrocketing data volumes, increasing complexity in the variety of data, and raised expectations of performance and speed of analytic response continue to tax traditional enterprise data warehouses (EDWs) built on a staple of relational databases and SQL. This has given rise to a new breed of analytic databases, engineered from the ground up to service analytics queries at both scale and speed. Ovum's research has shown that an increasing number of organizations are now revisiting their current analytics strategies, and are looking to implement specialized analytic databases that promise to keep pace by delivering more advanced analytics in bigger and faster data environments. However, choosing a database is not straightforward. Several alternative analytic database architectures have emerged, including massively parallel processing (MPP) databases, columnar databases, in-memory databases, NoSQL databases, and hybrids thereof.

Ovum has identified eight leading analytic database solutions (IBM PureData, Kognitio, Infobright, SAP HANA, SAP Sybase IQ, Teradata Aster Discovery Platform, ParAccel, and HP Vertica) that provide the widest range of capabilities for organizations. This report outlines the strengths, weaknesses, and best use cases for each solution, and offers a comparative framework for evaluating each.

Ovum view

The business case for purchasing and implementing an analytic database is becoming clearer. Traditional data warehousing architectures are struggling to manage larger data volumes, handle new types of human- and machine-generated data, deliver rapid query response, and devise more sophisticated analytics. While companies can always throw more hardware and processing resources at the problem, this is not a cost-effective approach for corporate IT departments that are being asked to service more analytic requests with less budget and fewer resources.

To address these newer requirements, vendors have incorporated some core design principles in both their software and hardware for improving analytic processing performance. Software strategies include conventional query optimization, smart index management, pre-calculated/aggregated views (multidimensional or materialized), in-memory processing, and columnar storage. Hardware strategies that help to speed and improve query performance include using architectures such as MPP, in-memory engines, interconnects, and pre-configured appliances. All vendors covered in this assessment use one or more of these strategies to reconcile analytic performance with increased data scale and complexity.

Our research finds that the use of parallelization is the most commonly applied strategy – all of the solutions we evaluated support parallel loading and parallel querying of data. Support for in-memory processing technology is also a common thread across the group. However, the level to which solutions use in-memory varies widely across vendor solutions. SAP (HANA) is the most aggressive on using memory for loading and calculations. Others, such as IBM PureData, insist that analysis of large data sets make in-memory cost-prohibitive for installations above 5–10TBs, and instead rely on pinning frequently used data in memory for easy retrieval (caching). Other hybrid strategies employ a "tiering" mechanism, in which the most relevant or hot data is cached in memory, while cold data is moved to disks. Ovum believes that all these strategies are useful in specific analytic scenarios: some are effective for organizations that need consistent realtime responses on very large data sets, while others are effective for organizations that believe in prioritizing some data sets (in terms of response time) above others. An ideal – for performance purposes – would of course be to put all data into memory. However, despite the falling cost of memory, it is still more expensive than partial disk and memory platforms, and the cost-versus-performance differential influences enterprises' selection processes.

Six databases covered in this report are either "hybrid" or purely column based. Two – namely IBM PureData and Kognitio – are row based. In general, columnar architectures are believed to be better for analytical purposes; however, IBM and Kognitio are mature products, and the respective vendors argue that ad-hoc queries on frequently changing data are better served with row-based platforms.

The analytic database market is not vertical specific. Few vendors covered in this report provide extensive content and models for specific verticals. Consequently, organizations deploying these solutions need to engage IT services providers to build customized models. Ovum believes that building vertical content and expertise is a function of solution maturity and strategic direction. SAP Sybase IQ, arguably the most mature solution in this group, has extensive vertical content, while newer vendors such as Infobright rely on partners to embed industry-domain knowledge into its technology or into existing vertical applications through OEM licensing agreements.

All databases covered in this report provide strong support for advanced analytics, either in-database, through UDFs (user-defined functions), or by supporting parallelized versions of R open source statistical software. Enterprises are advised to explore the depths of the predictive libraries of the vendors they shortlist – these can range from scores to thousands of pre-built functions. Enterprises must also be aware that exporting existing statistical models directly to a newly deployed analytic database might entail modification of existing code and additional expenditure on services.

Organizations that are used to performing their analytics on traditional databases might need to invest time and effort in refurbishing models and data, as familiar constructs such as indexes, views, and joins might not be available or required on most of the analytic databases we assess. While some databases such as SAP Sybase IQ and Teradata Aster support indexes (but do not necessarily need them), most vendors advise customers to start with no indexes as these constructs take up additional space and increase complexity. Almost all databases offer extensive support for SQL, up to SQL:2008.

Lastly, all vendors in this report support faster deployment through cloud and appliances. The approaches to cloud and appliances vary, with some such solutions (such as SAP HANA) available only on pre-configured appliances or private clouds, while others (such as Kognitio) are available as commodity hardware. Cloud deployment will undoubtedly be a key area of development for all database vendors in the medium term.

Key findings

- IBM PureData, SAP HANA, Teradata Aster Discovery Platform, and HP Vertica are market leaders in the 2013–14 Ovum Decision Matrix for analytic databases.
- Kognitio, Sybase IQ, and ParAccel are strong challengers that have the potential to break into the leaders category within the next year, while Infobright appears to have a strong foothold in cloud-based deployments.
- All vendors covered support parallelization in data loading and querying.
- In-memory support varies within the group, with SAP HANA being the most aggressive, supporting a purely in-memory architecture, while others rely on a mix of disk and in-memory.
- The vendors in this report support information lifecycle management and data tiering (support for hot and cold data) to varying degrees, with mature vendors offering more support and younger vendors relying on third parties.
- Vertically oriented analytic models and content remains a key area of development, with very few vendors currently offering industry-specific content, schemas, and configurations for key verticals, even for larger verticals such as telecoms and retail.

VENDOR SOLUTION SELECTION

Inclusion criteria

What constitutes an analytic database remains a point of contention because any data store can arguably be considered fair game for analysis. Before selecting an analytic database, it is important to set reference and categorization guidelines. Ovum categorizes and evaluates analytic databases

across three main database categories: advanced SQL databases (designed around MPP and columnar-based structures), in-memory databases, and Hadoop.

Standalone analytic databases must offer minimum functionality to be included in any of the three categories used in this report. They must offer:

- a method of data management and assimilation that is considered an improvement over traditional online analytical processing (OLAP)
- rapid analytic processing, using but not limited to hardware-sharing models (such as MPP), dedicated shared-nothing environments, columnar or hybrid row/column, and in-memory cache methods
- the ability to scale to at least 1TB implementations, even though the analytic functionality can be applied to smaller data sets
- delivery as an appliance deployment, either by itself or through third parties; as on-premise software; or in a private or public cloud
- some, or all, analytic front-end tooling, either directly or through pre-integrations with third-party tools vendors
- tools for developers and business users to easily build analytic applications on top of the database.

Exclusion criteria

Conversely, vendor solutions are excluded from this report if they match any of the following criteria:

- solutions that do not address the database or infrastructure layer
- conventional relational DBMSs that are not optimized for data analysis
- point-tools for OLAP analysis, reporting, visualization, and predictive analytics
- analytic solutions that have a significant portion of database functionality delivered through third-party products.

Vendor solutions

Ovum's taxonomy for categorizing different types of analytic databases is described below.

Advanced SQL databases

Advanced SQL database platforms fall into one or all of these categories:

- MPP databases, such as IBM Netezza, Teradata Aster, and Kognitio, which usually come in fixed-configuration appliances and are generally well suited for managing over 10-plus terabytes of user data.
- Columnar databases, which include HP Vertica, SAP Sybase IQ, Infobright, and ParAccel.
- In-memory databases, such as SAP HANA, which rely primarily on main memory (as opposed to disk) for computer data storage and analytic processing. They are faster than disk-optimized databases because the internal optimization algorithms are simpler and execute fewer CPU instructions, making them particularly relevant for data analytics.

However, HANA can be classified as a memory-centric database and supports ACID properties of atomicity, consistency, and isolation.

Oracle chose not to participate in the Ovum Decision Matrix. Pivotal's participation was limited.

Hadoop databases

Because of the unique characteristics and use cases, Ovum excludes Hadoop as a standalone analytic database. However, we do examine where Hadoop platforms support the evaluation criteria, where we expect future support, and analyze which criteria either do not apply to Hadoop or should be modified for the architecture and usage patterns endemic to Hadoop. Ovum intends to author a separate comparative report on Hadoop offerings in 2014.

Methodology

Technology assessment

In this assessment dimension, Ovum analysts develop a series of features and functionality that would provide differentiation among the leading analytic databases in the marketplace. The criteria groups identified for analytic databases are as follows:

Table 1: Technical evaluation dimensions and considerations

Technical evaluation dimensions	Key considerations
Data access and integration	Data is the lifeblood of any analytic database. Providing quick and easy access and updates to a range of data sources is critical for operational decision-making.
High-performance analytics	Analytic databases are designed for speed. Fast processing and rapid response are critical to ensure analysts stay focused and can iterate quickly through detailed queries on large data sets.
Scalability	Analytic databases need to scale in parallel to Big Data, allowing for advanced analytic operations against terabytes of data and the varied queries and analysis of large and diverse user communities.
Analytic extensibility	Analytic databases need to provide a flexible breadth and richness of analytic functions that can be easily extended to support sophisticated analysis.
Adaptability	The analytic system should be flexible enough to adapt to constantly changing data/data models and query/analysis requirements.
System manageability	Analytic databases should be designed to be simple to set up and administer, arguably more so than a relational database. The system should provide unified tools and interfaces that allow for easy administration of all aspects of the system.
System availability	The analytic database should perform reliably enough to support mission-critical business analytics, which mandates minimal downtime.
Deployment	Analytic databases should be rapidly deployable and fit comfortably in organizations' roadmaps and existing enterprise IT architectures, both on premise and/or in the cloud.

Source: Ovum

Execution

In this dimension, Ovum analysts review the capability of the solution around the following key areas:

- **Maturity:** The stage that the product/service is currently at in the maturity lifecycle is assessed here, relating to the maturity of the overall technology/service area.
- **Interoperability:** In this element we assess how easily the solution/service can be integrated into the organization's operations, relative to the demand for integration for the project.
- **Innovation:** Innovation can be a key differentiator in the value that an enterprise achieves from a software or services implementation, and this is assessed in this criteria.
- **Deployment:** Although not scored on the model, this refers to a combination of assessed criteria and points of information. Ovum analysts provide detail on various deployment issues, including time needed to deploy, industries served, and services and support offered.
- **Scalability:** Although not scored on the model, points of information are provided to show the scalability of the solution across different scenarios.
- **Enterprise fit:** Although not scored on the model, the alignment of the solution to different enterprises is assessed in this dimension, and the potential ROI period is identified.

Market impact

The global market impact of a solution is assessed in this dimension. Market impact is measured across three categories, each of which has a maximum score of 10.

- **Revenues:** Each solution's global analytic database revenues are calculated as a percentage of the market leader's. This percentage is then multiplied by a market maturity value and rounded to the nearest integer. Overall global revenue carries the highest weighting in the market impact dimension.
- **Revenue growth:** Each solution's revenue growth estimate for the next 12 months is calculated as a percentage of the growth rate of the fastest-growing solution in the market. The percentage is then multiplied by 10 and rounded to the nearest integer.
- **Geographical penetration:** Ovum determines each solution's revenues in three regions: the Americas; Europe, the Middle East, and Africa (EMEA); and Asia-Pacific. These revenues are calculated as a percentage of the market-leading solution's revenues in each region, multiplied by 10, and then rounded to the nearest integer. The solution's overall geographical reach score is the average of these three values.

Ovum ratings

- **Market leader:** This category represents the leading solutions that we believe are worthy of a place on most technology selection shortlists. The vendor has established a commanding market position with a product that is widely accepted as best-of-breed.
- **Market challenger:** The solutions in this category have good market positioning and are sold and marketed well. The products offer competitive functionality and a good price-performance proposition, and should be considered as part of the technology selection.

- **Market follower:** Solutions in this category are typically aimed at meeting the requirements of a particular kind of customer. As a tier-one offering, they should be explored as part of the technology selection.

Ovum Interactive Decision Matrix

To access the Interactive Decision Matrix for analytic databases – an online interactive tool that provides the technology features that Ovum believes are crucial differentiators for leading solutions in this area – please download the Ovum Interactive Decision Matrix tool from the Ovum Knowledge Center.

MARKET AND SOLUTION ANALYSIS

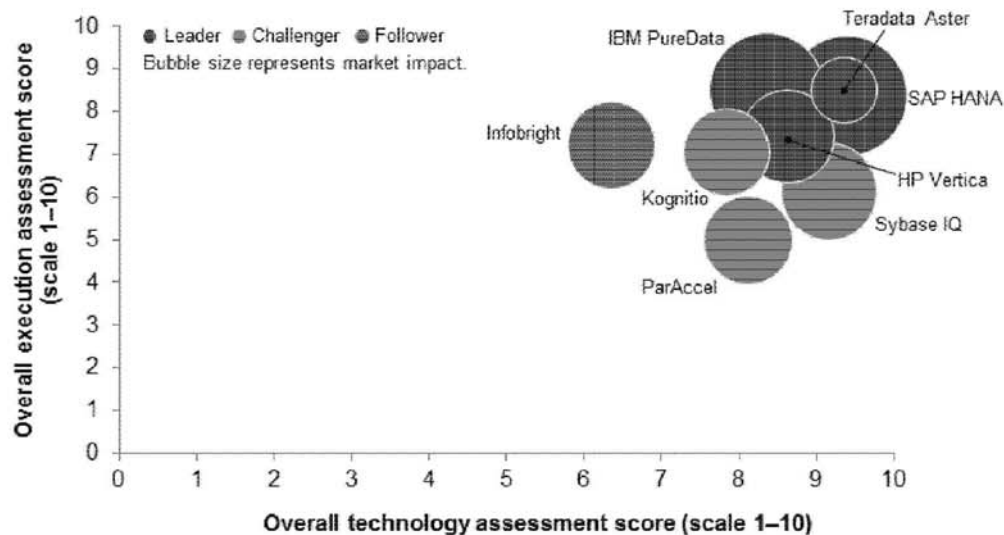
Ovum Decision Matrix: analytic databases, 2013–14

The aim of this report is not necessarily to provide an "apples-to-apples" comparison based solely on function. Rather it is intended to help organizations narrow down a shortlist of database platforms that match their business needs and existing IT infrastructure. Above all, our evaluation framework is not intended to replace a full in-house evaluation.

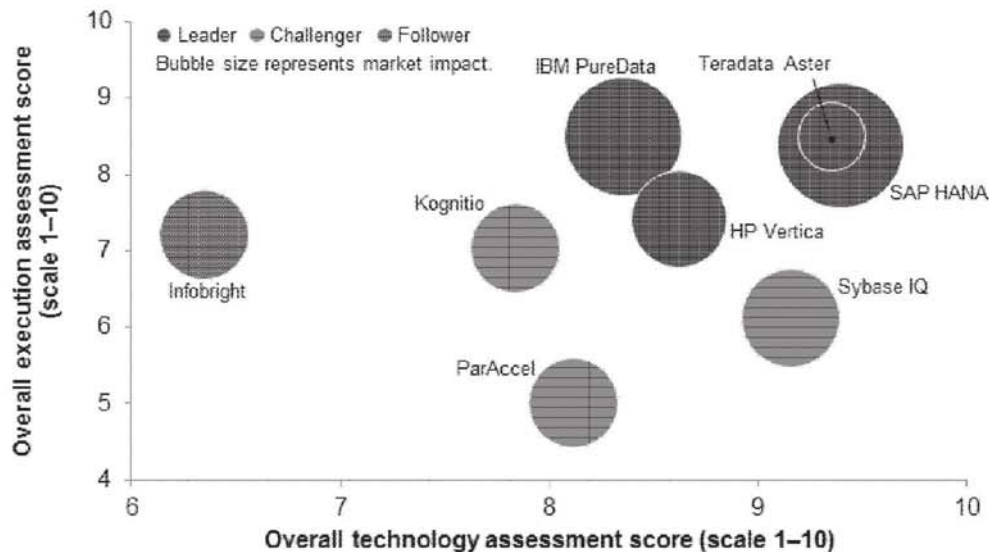
Oracle chose not to participate in the Ovum Decision Matrix. Pivotal's participation was limited.

Figure 1 shows the solutions clustered together in the full-scale Ovum Decision Matrix for technology (x-axis), execution (y-axis), and market impact (size of bubble), while Figure 2 zooms in on the upper right of Figure 1 to provide a closer look at vendor positions.

Figure 1: Ovum Decision Matrix – analytic databases, 2013–14



Source: Ovum

Figure 2: Expanded view of Ovum Decision Matrix – analytic databases, 2013–14

Source: Ovum

Table 2: Ovum Decision Matrix: analytic databases, 2013–14

Market leaders	Market challengers	Market followers
IBM PureData	Kognitio	Infobright
SAP HANA	Sybase IQ	
Teradata Aster	ParAccel	
HP Vertica		

Source: Ovum

Market leaders: Teradata Aster Discovery Platform, SAP HANA, IBM PureData, HP Vertica

The analytic database market is top-heavy, i.e. crowded at the top, with four players falling into our "market leaders" category. Of these, only SAP HANA is a homegrown solution, while the rest are acquisitions that have now been well integrated into the larger acquirer's portfolio.

IBM PureData (formerly Netezza) occupies the top spot, due to its technological prowess and market-leading execution capabilities. While it did not earn the top spot on technology – partly because IBM does not pitch PureData as an all-inclusive analytic option, but as one of three analytic database platforms designed for transactional, analytic, and operational workloads – it is extremely strong in its core area, as a platform-agnostic analytic database with minimal tuning and deployment concerns. However, what makes IBM PureData stand out from the rest of the pack are execution and the company's ability to customize content and expertise for all major verticals, which is a key differentiator in a crowded marketplace.

The Teradata Aster Discovery Platform also scores highly in our assessment – a testament to Teradata's long and rich heritage in supporting some of the largest EDW environments. To its credit, Teradata has also recognized the shift away from centralized data warehouses to more specialized appliances, as demonstrated by its broader portfolio of purpose-built analytic database offerings. The Teradata Aster Discovery Platform shines when used for multi-structured data exploration. The solution complements existing EDW environments well and is supported by Teradata's mainstream database management software and Infiniband connections under a tightly unified architecture. It scores the highest among the group on high-performance analytics and deployment, and we believe it is one of the most innovative solutions on the market today, with a secure place among market leaders. Teradata also stands out due to its focus on unification of different data-processing engines within the Teradata family and due to its management, coordination, and routing of analytics to the right engine. With the help of its Unified Data Architecture (UDA) and SNAP framework, Teradata can coordinate query processing between SQL, SQL-MR, and SQL-GR, and multiple data stores, such as row, column, and file store. At this point, no other analytics platform provider offers the same capability to manage queries across SQL and Hadoop/NoSQL engines.

SAP HANA is a well-rounded solution, with leading scores on execution, market impact, and technology. SAP HANA's differentiated architecture, with a strong focus on in-memory, sets it apart from other vendors. Also, SAP HANA's capabilities to double up as a transactional database capable of both OLTP- and OLAP-style operations makes it relevant to a larger number of use cases. While deployment in non-SAP shops is happening, the key reason for the high market impact of SAP HANA lies in its success in penetrating the vast SAP customer base.

HP Vertica joins the leaders group based on good scores on both technology and execution. Vertica scores highly in maturity and technology, offering a strong columnar store with data-tiering capabilities that help to further optimize the platform's performance and storage capabilities. The vendor ranks highly on execution due to its flexibility in deployment, a versatile community edition, and strong support and services. We believe that the vendor has a lot of unexplored potential around database management and additional portfolio integration. Should HP be successful in tightly integrating its stack, especially Autonomy, with HP Vertica, the vendor could well outdo its peers and come out at the forefront of structured-unstructured data analytics. However, much depends on whether HP continues to invest in Vertica to help it gain further momentum. This is a key point: Will HP Vertica thrive in HP's current disorganization? The underlying technology is not in question, but HP's executional woes might well hinder wider adoption.

Market challengers: Kognitio, ParAccel (now Actian), Sybase IQ

The challengers category comprises three databases of different lineage and backgrounds. Two vendors, Kognitio and ParAccel, are somewhat similar in size and extremely promising in terms of revenue growth and technology. However, being relatively small vendors in a crowded marketplace is difficult, especially when they need to compete with mega vendors head-on. Both these vendors have done remarkably well in specific geographies (UK) and scenarios (attractive price-performance ratios). However, both will need to expand their marketing and sales forces to become true global players.

ParAccel was acquired by Actian while this report was being authored and this could positively affect its competitive positioning in 2014. However, whether the acquisition is accretive or not will depend on how central to Actian's portfolio ParAccel is. Actian has been on a spate of complementary acquisitions over the past year, including Pervasive Software and Versant. Ovum believes that Actian's ParAccel MPPdb will appeal to organizations looking for a software-only solution (though it is available as an appliance) that offers attractive price performance and leverages disk or in-memory to overcome SQL's scale and performance limitations in complex analytic processing cases.

Kognitio is by no means a new company, although it has recently come into focus after recording near-100% growth in revenues. The product scores well on technology by pursuing a clear strategy of converging in-memory and cloud analytics to target near-realtime use cases. A key differentiator, in Ovum's view, is Kognitio's strategy of pinning data in memory persistently throughout an entire processing task while retaining synced tables/views. The vendor scores above average in data access and integration, high-performance analytics, adaptability, and deployment. If the company manages to keep growing at current rates, it will enter the leaders' category in the next two to three years. However, the vendor needs to expand its global footprint beyond its core UK market, particularly in North America, where tie-ups with more resellers and partners can help.

SAP Sybase IQ sits on the cusp of the leaders and challengers section. It is one of the most mature analytic database products in the market with enviable customer references in the financial services industry and a strong partner network for support and implementation. However, it is a product that is arguably past its prime – especially in regards to its reliance on symmetric multi-processing (SMP). Ovum finds that Sybase IQ competes with and surpasses many vendors' offerings in this assessment today, due mainly to SAP's continuous efforts to retrofit newer features into the database. It will certainly appeal to "conservative" IT organizations that prefer a proven solution with solid references. The solution offers best-of-breed technology in data access and integration, scalability, and system management. SAP continues to support the database and is introducing it to a large number of its own accounts in bundled deals with SAP HANA. However, Ovum believes that Sybase IQ runs the risk of being overshadowed by SAP HANA over the longer term; this is part of the reason why this product is placed in the challengers category.

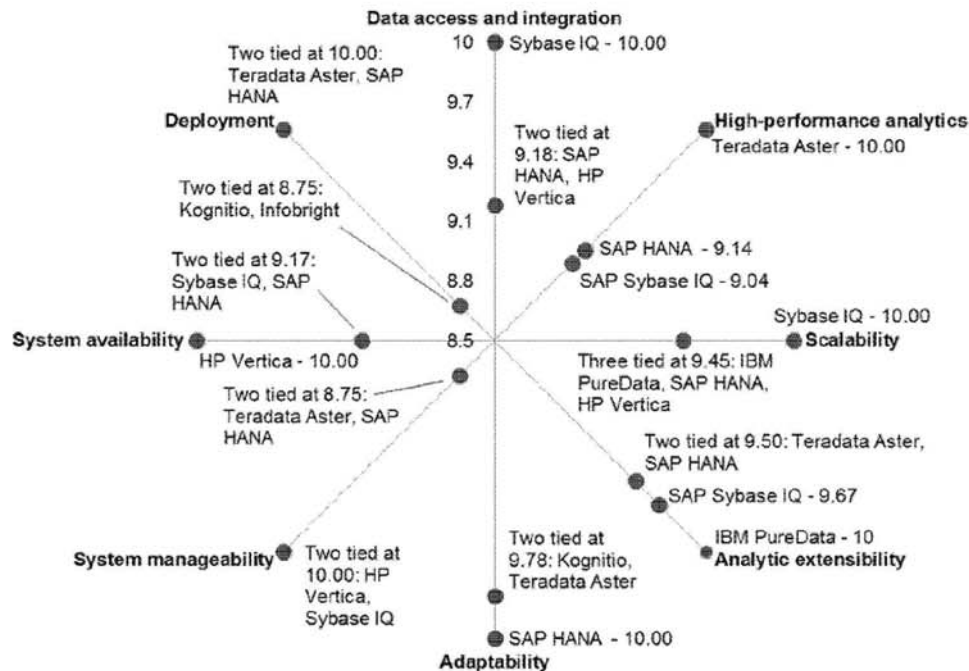
Market followers: Infobright

Infobright offers a columnar, highly compressed database that is built on open source database technology and is optimized for accurate and rapid ad-hoc analysis of large volumes of machine-generated data. While somewhat of a niche analytics player, Infobright has several technically differentiated features that have resulted in a steady widening of its customer base for the community and enterprise editions of its software and its new database appliance. The company has intelligently targeted leadership in a small (but increasingly important) analytic target – namely, machine-generated data. However, it could likely be considered niche by organizations that are looking for more rounded capabilities across multiple data types.

MARKET LEADERS

Market leaders: technology

Figure 3: Ovum Decision Matrix: analytic databases, 2013-14 – Market leaders – technology



Source: Ovum

The market leaders diagram (Figure 3) depicts the leading vendors in each of Ovum's technology assessment categories. We typically present three leading vendors for each category. However, if multiple vendors receive the same score, then all vendors sharing the same score are displayed.

For this assessment, we chose eight broad technology parameters to evaluate the state of technology evolution and the maturity of the analytic databases market. The results show that while the market is nowhere near maturity, it is no longer in a nascent stage either. Vendors are slowly veering towards a common set of "must-have" features and best practices, which is most apparent in categories such as system manageability, system availability, and deployment, where the variance between vendor scores was minimal. Almost all vendors agree about the importance of security and auditing tools for database administration, provision of ample redundancy, and the need to provide interruption-free disaster recovery and fault-tolerance features. The maximum variance in scores was in analytic extensibility and high-performance analytics, where some vendors offer a well-rounded set of generic capabilities, while others are purpose built for a specific use case.

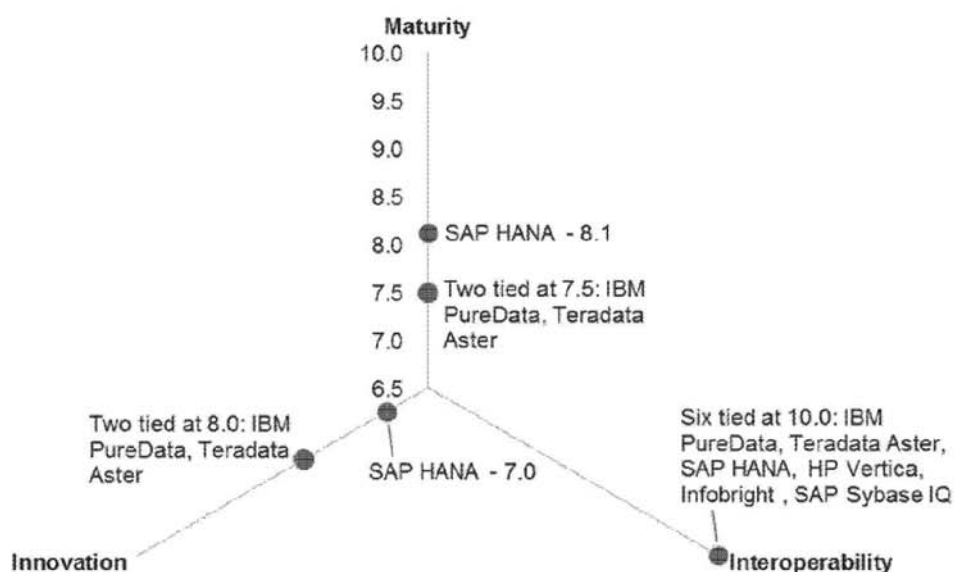
Teradata Aster Discovery Platform Database and SAP HANA lead this assessment with maximum average technology scores. They are followed closely by Sybase IQ, HP Vertica, and IBM PureData. Teradata Aster Discovery Platform Database scores above average in all technology categories, leading the group on high-performance analytics and deployment. Ovum finds Teradata's philosophy of

implementing modern architectures while allowing users the freedom to maintain old constructs (indexes, materialized views) a step in the right direction. Also, Teradata Aster Discovery Platform's SQL-MapReduce stands out as a practical way for enterprises to bridge the SQL and NoSQL worlds in an analytic context.

SAP HANA also scores above average in all categories. Its unique architecture and focus on in-memory, which is particularly advantageous in increasing query and data-loading speeds and in reducing the need for pre-processing of data, sets the product apart from other vendors' offerings.

Market leaders: execution

Figure 4: Ovum Decision Matrix: analytic databases 2013-14 – Market leaders – execution



Source: Ovum

Enterprise data warehouse deployments were traditionally considered difficult and complex, requiring services for all but the simplest of deployments. Analytic databases evaluated in this report alleviate this problem partly by being rapidly deployable (through appliances) and fitting readily into existing enterprise architectures – both on premise and in the cloud. However, migrating from one platform to another or from one version to the next is not without its difficulties. Most vendors offer a toolset to help with the migration, but in most cases additional services-led support may be required. In terms of implementation time, a modest deployment with a few users can be achieved quickly. But large complex implementations invariably take several months, regardless of which vendor's platform is selected.

All analytical databases we evaluated had very good support for upward scalability, meaning that adding more hardware and licenses nearly always led to linear scale-up. However, actual deployment numbers told a different story. Customers rarely provided analytical database access to the entire

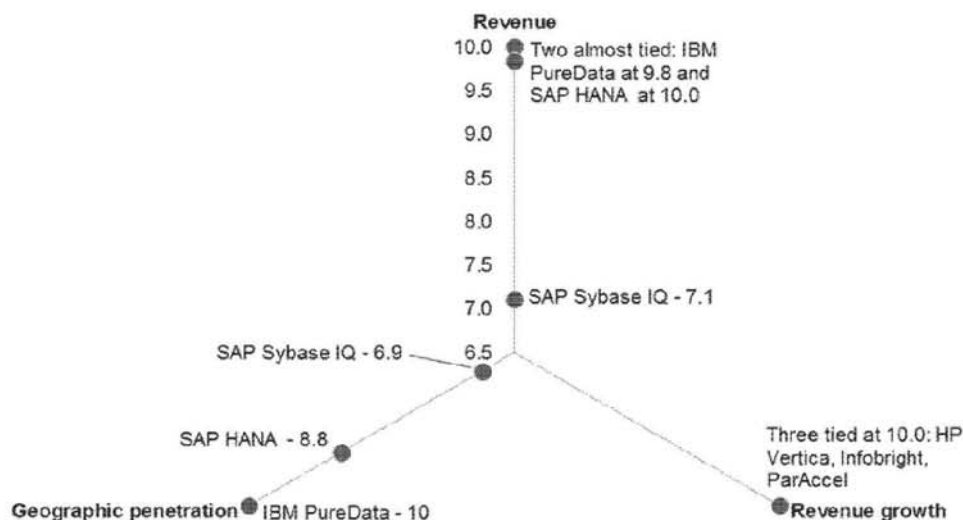
enterprise, instead limiting it to power users and analysts. That in turn limited the actual number of users of analytical databases to the low hundreds.

Overall, Teradata, IBM, and SAP HANA lead the way in execution. All of these vendors have very strong global sales teams that are supplemented by local resellers and vertical specialists, who help in customizing the solution to verticals and geographies. Both Asterdata and Netezza have been able to use Teradata's and IBM's distribution networks and brand effectively to build up a global presence.

In terms of maturity, SAP HANA, IBM PureData, and Teradata Aster Discovery Platform share the leading spot. IBM and Teradata both have many years of experience in databases, while SAP HANA possesses many differentiators architecturally and is gaining rapid traction the marketplace. Interoperability sees almost all vendors tied for the highest score, with only two vendors relying on add-on or third-party products for completing data integrations. Widespread support for open standards has made interoperability between databases much easier, allowing organizations to adopt a best-of-breed approach. Most analytic database vendors also provide integrations with line-of-business (LOB) applications and BI products. In terms of product and marketing innovation, Teradata and IBM stand out most, though SAP HANA follows closely.

Market leaders: market impact

Figure 5: Ovum Decision Matrix: analytic databases, 2013-14 – Market leaders – market impact



Source: Ovum

Market impact is an important part of our assessment as it reflects how effectively each vendor has penetrated different geographical areas and various vertical markets.

In terms of geographical penetration, IBM (not surprisingly, given its size and global reach) scored the highest, with SAP HANA being a distant second. SAP Sybase IQ followed SAP HANA closely, finishing

third. Overall, IBM and SAP HANA come out at the top due to their large revenue base, high growth rates, and equitable geographical distribution.

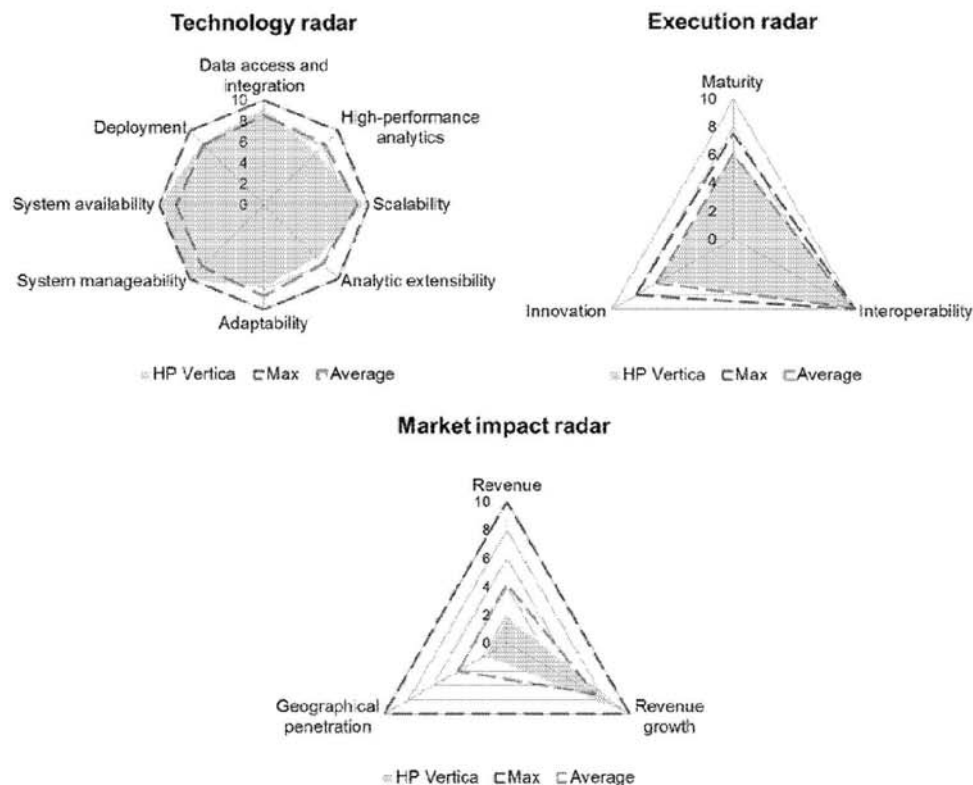
Ovum has estimated vendor revenues and growth rates as part of the metrics in this section. Given the attractiveness of this sector, heavy marketing budgets, and the smaller size of some companies, average growth rates were approximately 70% in 2012. However, this figure should be taken with a pinch of salt; smaller vendors in this assessment with less than \$50m in revenues recorded growth rates of around 100%, which skewed the average upwards. Among the larger vendors, SAP HANA was estimated to grow at the highest rate, mainly due to its use in both analytical and transactional use cases for existing SAP applications. Over the next two years we anticipate many of the vendors featured in this Ovum Decision Matrix to record low double-digit growth rates as the market stabilizes.

Interestingly, smaller vendors such as Infobright, Kognitio, and ParAccel recorded growth rates of almost 100% in 2012. Of this group, ParAccel was acquired by Actian in April 2013, and Ovum believes that it might not be the last one. Infobright and Kognitio remain attractive acquisition targets, and could be acquired by larger vendors over the next two to three years, not necessarily for their customer bases, but as technological differentiators.

VENDOR ANALYSIS

HP Vertica (Ovum recommendation: Leader)

Figure 6: HP Vertica radar diagrams



Source: Ovum

Ovum SWOT assessment

Strengths

Optimal price-performance for specific analytic use cases – The Vertica system is built to scale on a shared-nothing deployment, using commodity MPP clusters, and its separation of an in-memory write-optimized row store from a read-optimized disk-based column store improves load speed (which has always been a problem for columnar databases). Vertica delivers "fast-performing SQL" to ensure performance, particularly for self-joins used in graph analysis – mirroring the needs of its numerous telco and Internet company customers that extensively use graph analysis. Vertica also offers a library of in-database analytics, including parallelized functions for predictive algorithms based on R. Price-performance value is aided by Vertica's pricing model, which is based on the amount of data

extracted from source systems and loaded into the database – rather than the number of users, servers, chips, or processor cores.

Strong columnar design traits – As a columnar-designed analytic database, Vertica has a natural advantage in columnar compression-ratio capabilities. Depending on data type, Vertica can compress data up to 70–80% of its original size, even while maintaining high availability, redundancy, and querying speeds. While this is no longer a unique technical differentiator, Ovum finds that Vertica's compression compares very favorably to peers' row-based systems and pure in-memory systems.

Intelligent tiered storage for hot and cold data – HP Vertica also offers three different levels of storage that allow users to easily prioritize storage-tier locations and define policies for storing data objects on these tiers. A key feature is storing data in the way it is most likely to be queried – rather than simply storing data in tables and adding secondary structures, such as indexes and materialized views, for querying, that result in redundancy, an artificial increase in database size, and added DBA (database administration) overheads. Vertica stores all data in its physical storage in optimized collections of table columns ("projections"), based on dynamic analysis of the frequency by which it is accessed and queried, i.e. hot, cold, and offline. Vertica creates these projections automatically, during (or shortly after) data loads and can also be optionally stored in "overlapping" views, each of which is sorted by at least one of the columns in the view. The projections are abstracted from the user so that standard SQL can be used against the database and the Vertica query optimizer automatically picks the best projections to use for any query. Although these projections come with a level of redundancy, the net benefits are faster querying, auto refresh, and dynamic changes, and are especially useful for deployments that have access to a range of storage devices such as Flash, HDDs, and SANs. A significant feature is the archiving of partitions to offline storage; unlike traditional data backups, partition archiving moves selected partitions out of HP Vertica online storage rather than creating a backup copy of the data.

Flexible deployment options – Vertica is a flexible analytic database, offering customers a range of choices in deployment models and software/hardware platforms (excluding Windows). The vendor offers its analytic database in both pure software models (to be deployed on industry-standard x86 servers) and as an appliance (rolled out post-merger and integrated with HP ProLiant DL380 servers), which is purported to support up to 1,000 users on a single appliance rack. Additionally, Vertica can be deployed on private clouds, using virtualization infrastructures such as vSphere from VMware and on public clouds via HP Cloud Services and Amazon's Elastic Compute Cloud – thereby availing the dual benefits of virtualized consolidation in the cloud and potentially providing easier management, high availability, data-center energy savings, and faster test to production. Vertica can also be deployed on-demand, as a hosted, managed service.

Versatile community edition – While other database vendors offer similar community editions of their databases, which allow organizations (and academia) to test and trial the database without actual purchase, Ovum finds that Vertica's community edition offers the highest level of functionality, with access to almost all the paid-version features, except support. The community edition does not have any cost or time limit, allows up to 1TB of data storage across three nodes for an unlimited time, and provides access to a community for downloads and documentation. This flexibility will endear Vertica to academia and data science students, while giving organizations the ability to run large test-cases on their own terms before purchasing the solution.

Weaknesses

A few missing database functions – While Vertica's performance is not in question, the platform could be enriched in other ways, particularly around management features like integrated workload management, geospatial support, and better support for APIs. We feel that HP has not yet fully exploited its leadership in system management tooling, and has yet to leverage many of these capabilities into the core Vertica environment. HP assures us that it will soon initiate support for geospatial data. Ovum would also like to see HP Vertica provide better support for a range of APIs, including Pro*C/C++ Precompile and ADO.NET, and royalty-free APIs. HP must act quickly to keep pace with increasingly complex and concurrent analytic workloads and demands.

Piecemeal integration with rest of HP's information management portfolio – HP has been slower than the industry in using HP's information management stack to create vertical- and function-specific applications. HP continues to lack a comprehensive integration roadmap for the multiple products in its portfolio, which largely remain under separate management within the organization. For Vertica, that might prove to be a costly mistake over the longer term when database performance will no longer be a key selling point. One area that cries out for clarification and deeper integration work is with Autonomy's Intelligent Data Operating Layer (IDOL) platform (which HP acquired back in 2011), in order to blend complementary structured and unstructured data querying, analysis, and management. There are positive signs, however; the new relationship between Autonomy and Vertica, which sees peer-level general managers reporting to the executive vice president of HP Software, seems to have stabilized matters. In June 2013, HP announced its HAVEn initiative to combine products from Autonomy, HP Vertica, HP ArcSight, and HP Operations Management with Hadoop to create a more broadly functional Big Data analytics platform; however, the first solution to be built (HP Operations Analytics) has focused solely on optimizing internal IT operations.

Lack of specific technical or business differentiators – In such a rapidly developing market, Vertica's column-store and MPP architecture are no longer differentiators. Vertica will face stiff competition from larger and more mature analytic database vendors that are also aggressively pursuing appliance-driven strategies. Vertica will also be pushed hard from vendors such as Kognitio and Actian that focus on commodity hardware deployment. At this stage, most database vendors have added MPP architectures and columnar or hybrid compression capabilities to their products. To survive in the longer term, Vertica will need to take its pick and specialize in a few analytic use cases rather than play a jack-of-all-trades role. It will also need to better use HP's marketing and supplier muscle to promote a refined value proposition to clients.

Lack of verticalization – Admittedly, Vertica does not have a strong vertical focus, which both HP and Autonomy lack as well. Ovum believes that as the market for analytic databases matures, performance improvement will become "table stakes" and vendors will increasingly need to differentiate on function and vertical-specific solutions – either packaged or as a consulting engagement. Currently, HP Vertica is focusing on the financial services, healthcare, and communication services segment. Ovum believes HP has to invest in creating data models and content for verticals that could help in reducing time-to-value for its customers – an area where HP can improve its professional services capabilities to create productized analytic services on top of the HP Vertica Analytics Platform. One area perhaps

relevant for financial services is investing in CEP (complex event processing) technologies, an area in which financial services firms have been early adopters for applications around rapid-fire stock trading.

Sluggish sales momentum – HP Vertica's sales momentum has disappointed, despite early successes as an independent company. The company has also been slow to exploit global presence. In 2012, Vertica's customer base was still overwhelmingly in North America, a reflection of HP's slowness to capitalize on its worldwide sales and marketing reach. However, new sales force training and more attractive compensation arrangements seem to be accelerating global adoption.

Opportunities

Piggyback on HP's cloud initiatives to target SMBs – Historically, the lion's share of HP Vertica's business came from software or virtual machines. With HP's hardware and cloud infrastructure expertise, Vertica can target a strong base of cloud-based clients, focused on HP's public and private cloud solutions. Vertica has been one of the first vendors to offer a cloud instance in Amazon, but has not been able to commercialize its database cloud. HP's persistent push towards a cloud-based strategy should continue to open up further opportunities for Vertica, particularly in the highly price-conscious SMB segment.

Delivering an end-to-end structured-unstructured data analytics platform – While Vertica performs well on structured data, it has been slow to devise an integration strategy that combines its skills with unstructured data (Autonomy IDOL), archiving, and NoSQL analytics (Hadoop). HP is starting to address this with its HAVEn initiative, which Ovum believes is a good direction for the company to pursue given that no other analytic database vendor has expertise in dealing with both structured and unstructured data in the same software solution. While the lack of integration with Autonomy is not such a big deal right now – as most analytic database vendors do not own a text search engine – it remains a missed opportunity for HP and Vertica to differentiate itself in the market. HP certainly has the hardware expertise to undertake this project, but with higher customization and integration it could risk making Vertica less stack-neutral. Another opportunity is to tap into HP's hardware and system management resources – such as HP's Vertica Management Console and Cluster Management Utility (CMU) – to provide value-added Web-based management, configuration, and monitoring functions for 3-D cluster visualization, time-series performance and diagnostic graphics, alerting, and "push-button" cluster operations.

Leverage partnerships through alternative delivery models – HP now claims over 2,250 direct, OEM, and channel customers for Vertica. Vertica's SaaS and embedded capabilities are certainly attractive to HP's broad ecosystem of partners, notably services providers and large system integrators like Accenture and Capgemini, as well as smaller vertically focused shops. HP also partners with a range of ISVs who are creating products using Vertica. HP Enterprise Services also now provides data models and also supports the MyVertica community's participation in open source code sharing.

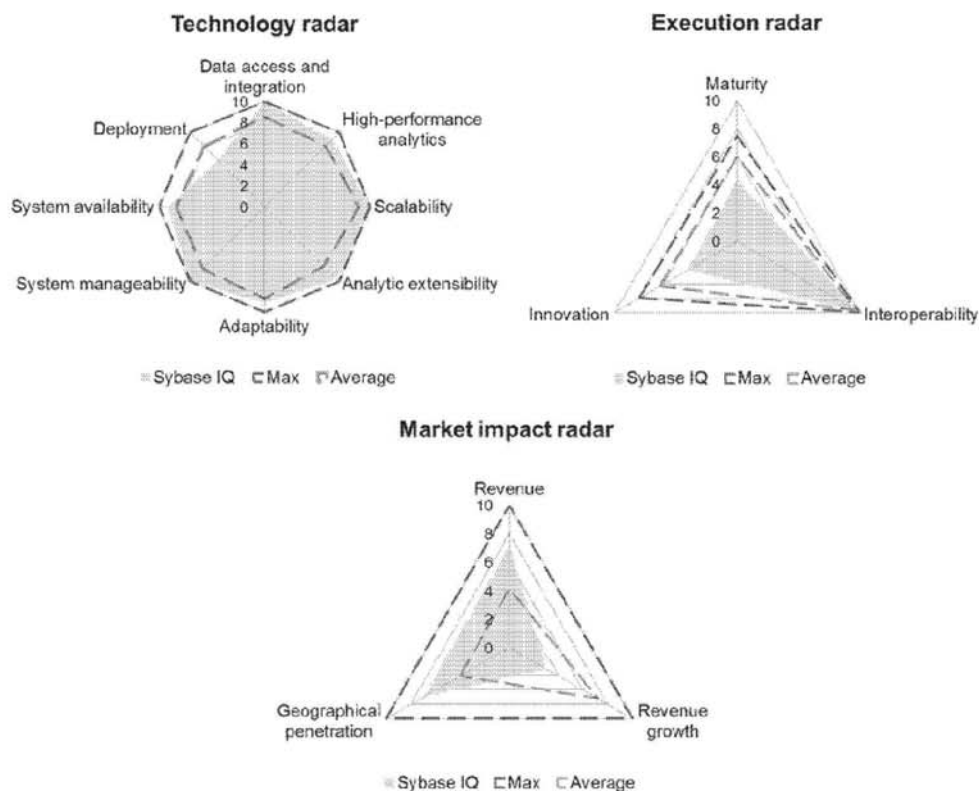
Threats

Negative customer sentiment and market confidence – HP's acquisition of Vertica has not yet translated into clear synergies for both companies – evident from lackluster sales growth in Vertica post-acquisition. That puts into question HP's ability to run Vertica either as a standalone or integrated

part of its overall business. HP has had a checkered past of effectively managing technology acquisitions (related to the acquisitions of EDS and Autonomy) and recent accounting scandals and write-offs around Autonomy hardly muster up confidence among both customers and shareholders. Arguably, HP's troubles are less relevant to Vertica. But many customers still view HP as a systems and hardware company that has little in common with Vertica, reflective of HP's limited experience in analytics services, and one that might not be able to ensure Vertica's long-term growth and viability.

SAP Sybase IQ (Ovum recommendation: Challenger)

Figure 7: SAP Sybase IQ radar diagrams



Source: Ovum

Ovum SWOT assessment

Mature columnar analytic database with loyal customer base – Sybase IQ is one of the most mature columnar databases in the market, with over 2,000 customers and two decades of R&D investment behind it. SAP boasts hundreds of reference customers and case studies for IQ Server, showcasing its success in a variety of IT environments, especially legacy banking and financial services industries.

This is matched by a strong ecosystem of complementary offerings, which position SAP to provide a full portfolio of services around data warehousing.

The technical maturity is evident in Sybase IQ Server's data-modeling tooling (based on Sybase PowerDesigner), data integration tools (Sybase Replication Server and SAP Data Services), a complex event processing (CEP) engine called the Sybase Event Stream Processor (ESP), and text analytics and business intelligence (SAP BusinessObjects).

While the product has been sold to customers across industries, SAP Sybase IQ is most well known in capital markets and financial services sectors that naturally require realtime analytics. Sybase also combines Sybase IQ with complementary products to exploit specific use cases in the financial services industry: for example, Sybase's RAP platform, which is marketed heavily into the capital markets, combines Sybase IQ with an in-memory version of Sybase ASE and Sybase ESP.

Built from the ground up for high performance and scale – SAP Sybase IQ has been tested to scale for small (fewer than 100 users) to very large enterprise deployments (more than 5,000 users). In SAP Sybase IQ v16, the vendor has added a "petabyte-scale" store which further enhances scale-out and scale-up. The database, which was originally designed on shared-everything principles, can now use hash partitioning for query scale-out, which helps the query engine scale up and/or out more effectively. Other enhancements include a fully parallelized bulk loading feature, which is useful for loading large data chunks (volume) to the system.

Customer references that corroborate Ovum's assessment of SAP Sybase IQ's high scaling and performance capabilities include:

- **comScore Inc:** The company has a 147TB data warehouse, with 150GB of data loaded daily. SAP Sybase IQ provides marketing intelligence and analytics on this data, with support for granular data reporting with reduced query time.
- **Yapi Kredi:** The company has a 70TB data warehouse that is required to access enterprise data and perform complex ad hoc queries rapidly and easily. SAP Sybase IQ helped the customer increase the volume and effectiveness of customer marketing campaigns.

Superior compression capabilities – SAP Sybase IQ is a mature columnar database and has strong columnar compression capabilities. Depending on data type, Sybase IQ can compress data up to 60–70% of its original size, even while maintaining high availability, redundancy, and querying speeds. While this is no longer a unique technical differentiator, SAP Sybase IQ's compression capabilities compare very well to pure row-based systems and pure in-memory systems. The product allows queries to run directly on compressed data without the need for decompression, which improves query time. Given its maturity, the product has performed well in several benchmarking assessments and recently achieved an audited result of 34.3TB per hour for loading and indexing unstructured data.

Flexible deployment and interoperability – SAP Sybase IQ supports a broad range of software and hardware platforms and applications, ranging from on-premise (with support for commodity hardware from multiple vendors and running on Linux, Unix, or Windows) to cloud. This approach is a contrast to other vendors that only run on "certified" hardware, including SAP HANA. SAP Sybase IQ customers can also, in contrast with SAP HANA, fully virtualize an "elastic" database cluster composed of modular compute and storage components that are assembled on the fly to cater to specific client needs. SAP

Sybase IQ has the ability to offer mix-and-match appliances with components from multiple vendors. Examples include the Sybase Analytic Appliance which comprises SAP Sybase IQ running on IBM Power Systems servers, Sybase PowerDesigner, and BI software from MicroStrategy. SAP Sybase IQ supports a hybrid of in-memory/caching and disk-based storage/execution.

Weaknesses

Lacks customizations for SAP portfolio – SAP has chosen SAP HANA as the backbone for powering SAP applications; SAP Sybase IQ is either pitched as a near-line store for SAP HANA, or for customers that do not intend to operate on the SAP BW (Business Warehouse) environment. While this does serve to demarcate SAP HANA and Sybase IQ's territory, it also means that SAP Sybase IQ will not be able to target SAP's large native customer base as a standalone analytic database. Ovum believes that this is an unfortunate result of Sybase's acquisition by SAP and could turn out to be a material "opportunity cost." In other words, SAP's acquisition of Sybase prevents the product from being sold into lucrative SAP shops as a standalone database, which potentially limits its target revenue base and results in lower revenues/sales than it could have generated had it been an independent company. However, SAP specifies that it is seeing a lot of traction for SAP Sybase IQ as a near-line store to SAP Netweaver BW at the moment.

Aged product originally architected for SMP and progressively retrofitted – SAP Sybase IQ was one of the first columnar analytic databases in the market and has a legacy customer base in the financial services industry. However, the product was built during the mid-nineties when MPP technology was in its infancy and symmetric multi-processing (SMP) was the popular architecture to achieve scale. While the SMP-versus-MPP debate remains largely academic, it is now widely agreed that SMP performs better with custom hardware and is more expensive to scale, while MPP is the option that most database vendors have adopted. Sybase's architecture of "shared everything" was originally built to cater to the SMP philosophy. The vendor has reengineered the product significantly after version 15.3 to introduce PlexQ, which adopts a hybrid approach to accelerate queries by distributing work to a grid. However, the approach adds additional query management and dynamic balancing overhead, although the difference may not be noticeable for most clients as many distributed queries scale almost linearly across a multiplex grid.

Opportunities

Target heterogeneous (non-SAP) environments – SAP Sybase IQ is well positioned to cater to non-SAP enterprises that are looking for an analytical data warehouse. Customers in this segment are usually looking to upgrade from a traditional RDBMS (for greater speed and data volumes); Ovum believes that SAP Sybase IQ, with its strong columnar compression, parallelization, and in-database functionality, offers a viable alternative to row-based RDBMS, at a lower price than fully memory-based systems. SAP Sybase IQ offers a mature product that uses standard SQL and therefore can be set up and run with readily available IT expertise.

Become a staging/middleware layer for Hadoop at SAP – Since version 15.4, Sybase IQ has expanded its user-defined functions (UDFs) for accommodating native MapReduce functions that run within Sybase IQ and are invoked with SQL. This means the database has considerable potential to integrate Hadoop results into an SAP Sybase IQ database in several ways, including ETL processing

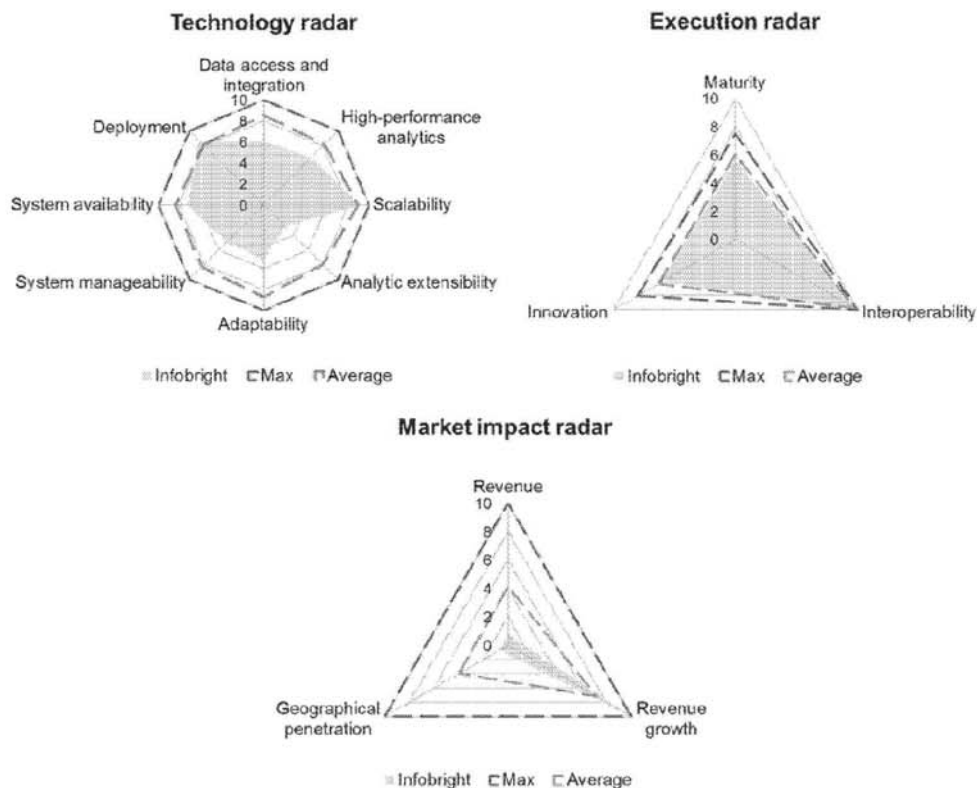
using open source SCOOP, data federation, query federation, and client-side federation using the TOAD SQL tool.

Ovum believes that SAP Sybase IQ, with further integration into Hadoop and HDFS (Hadoop Distributed File System), can take on the role of an offline analytics layer sitting between Hadoop and SAP applications. In such an arrangement, organizations could use Hadoop at the bottom layer for ingesting big, unstructured data sets, using Hadoop to economically run pattern recognition/indexing algorithms for sifting through Big Data. The result/reduced data set, which by then would be reasonably structured, could then be fed into Sybase IQ for analytics with familiar SQL tools in a relational MPP environment.

Threats

Hadoop maturing to an extent where it can serve as reliable near-line storage – To customers that already have a significant SAP investment (SAP BW, SAP Business Suite), SAP positions SAP Sybase IQ as the bulk/Big Data store for SAP HANA. This approach works well given that SAP Sybase IQ is a disk, cum memory-based, system which naturally works out to be cheaper than a complete in-memory system such as SAP HANA, especially for cold data.

One of the key emerging uses of Hadoop is as an inexpensive scalable near-line analytics data store. Ovum believes that the emergence of Hadoop tooling will expedite this process and as a result SAP Sybase IQ could be forced to compete against Hadoop for retaining its near-line store status.

Infobright (Ovum recommendation: Follower)**Figure 8: Infobright radar diagrams**

Source: Ovum

Ovum SWOT assessment**Strengths**

Analysis of machine-generated data – Infobright has continued to focus on an operational and investigative analytics niche around machine-generated data. Its Knowledge Grid technology clearly excels with machine-generated data and offers strong compression of proliferating amounts of streaming data that are stored and analyzed later – making it a good choice for web-generated data (clickstream data from Internet interactions), customer data records in the telecoms space (where it has had strong success with partners), and sensor data, like smart meter data in the utilities space. As machine-generated data continues to grow faster than other data types, and as companies increasingly learn how to leverage this data in operational analytic applications like network optimization, security and fraud, and online ad/content, Infobright's sharp focus on machine data will help the company to differentiate itself against larger vendors.

Well-designed metadata layer – Central to Infobright's capabilities is how it approaches data and brings it into its analytic engine. This is enabled using a highly differentiated, and powerful, metadata layer called Knowledge Grid, which is held in memory as much as possible (contingent to available system resources) for fast analytics. The Knowledge Grid automatically creates and maintains information about the data – including statistics like min/max, averages, deltas, and cardinality that are inferred from the data itself as it is being loaded and compressed into data packs – and uses this metadata for query optimization. Infobright employs its own portfolio of algorithms, based on its years of experience in developing business rules for handling the various data types commonly found in the machine-generated universe. This is akin to an indexing scheme that limits retrieval only to the data needed to resolve the specific query in question and to represent a first line of "analytic attack" for end users. The net result is a significant improvement in query performance (discussed below) and simplified database management. Storing "compact" information about the content and the relationships between compressed data packs and analyzing queries minimizes the need for data to be decompressed and removes the need to create special schemas, indexes, and aggregates, or perform manual performance tuning. These capabilities will hold great interest to OEMs looking to embed an analytic database that requires minimal back-end DBA (database administration) management and support.

Superior approach to compression and decompression – Delivering both strong compression and eliminating unnecessary decompression without conflict is one of the product's strongest differentiators, which stems from its metadata and data domain-specific approach. When data is loaded into tables, it is decomposed, and then further decomposed into separate data packs for each of the columns – broken down into the groups of 65,536 row elements – which are compressed as they are stored. By breaking each column by the same number of rows, it maintains its integrity with other columns for the same entry. Typically each data pack has a 20:1 compression rate, but this rate is often much higher; Infobright claims it is working to achieve up to 50:1 data compression rates in the future. Infobright tries to avoid decompressing the stored data (since decompression takes time and consumes valuable resources). Rather it takes its cue from the progressive refinement of query and analysis at hand; once end users hone in on specific domains of data to be queried, decompression occurs – in Infobright's parlance, from "rough" to "exact" calculation. Decompressing data this way is much faster than reading full-volume data from disk, and further enhances performance by limiting decompression to the data needed. Also the Knowledge Grid, which sits atop the data packs, adds an additional set of metadata and enables even greater performance.

Query performance – Infobright ICE and IEE are both designed as a low-touch analytical database and use its Knowledge Grid and theories of "granular" querying to leverage the native MySQL query cache and optimizer to enhance query performance. Accessing the Knowledge Grid to determine only those relevant data packs that apply to a particular query effectively rules out large parts of the Infobright database and significantly speeds up query processing and expedites the ad hoc discovery process. Since the corresponding data packs referenced in the Knowledge Grid are stored separately, an entire array of nodes can be queried without "disturbing" those data packs. One immediate benefit of this separation is that lock times are minimized, which increases query speeds. It also helps to categorize which data packs need to be decompressed through Infobright's Granular Query engine, which quickly isolates only the appropriate (and smaller) data sets, which are held in memory. User can

then drill down into these data sets for more-complex queries. To speed up the query experience even further, Infobright has introduced a feature called "DomainExpert" that provides specific intelligence about particular domains (URLs, IP addresses, and email address for web data). However, Infobright's approach only performs well for certain queries, particularly planned and repetitive ad hoc queries. Customers with highly diverse mixed workloads against a broader set of non-machine-generated data that requires additional preparation, updates, and analysis might need to reconsider Infobright's capabilities as they are not optimized for these uses very well yet, and some are not supported at all. Unsupported queries currently fall back to the standard MySQL optimizer, which do not work particularly well with Infobright's table structure. However, support for Postgres will handle them much better.

Rapid-fire data loading – A key innovation in Infobright is its Distributed Load Processor (DLP) utility, available only in the IEE version or in the Infopliance (which has its own version, called the Infopliance Data Processor), that allows for high-speed, parallel loading of compressed data into the Knowledge Grid's data packs on remote servers – at a purported load speed rate of one terabyte per hour. As Infobright often works in conjunction with extract, transform, and load (ETL) tools and platforms like Informatica and IBM, this remote data-loading capability can potentially be used to offload heavy ETL processing to a separate server and ensure more CPU resources are available on the database server for query processing. Separating Infobright's data inference and compression computations from the load process is particularly useful for ingesting bulky machine-generated data. Infobright also provides a Hadoop HDFS cluster connector with DLP, which many customers are increasingly starting to use for Big Data analytics.

Price-performance – Price-performance at a low cost is one of Infobright's strengths. Infobright remains one of the few open source column-store DBMS vendors with both open source (ICE) under a general public license (GPL) and subscription-based commercial (IEE) offerings based on the amount of raw data (SSED) stored in the system. The financial investment for an Infobright implementation is predicated on data size, not users. Solution configurations start at less than \$25,000 and can scale to more than \$1m. The open source cost model for ICE (no license fees) and a low-cost model for IEE (based on the amount of SSED), therefore makes it very interesting for organizations wanting to optimize traditionally expensive enterprise data warehousing deployments.

Weaknesses

Lightweight in-database analytics – Infobright lacks prebuilt analytic functions that can be executed in-database compared to other analytic databases. We find it surprising that the company does not provide prebuilt analytic functions for machine-generated data analytic scenarios – perhaps reflecting the company's lack of vertical partners. The potential extensibility and flexibility of analytic functions beyond machine-generated analytics is further hampered by the lack of an executable framework for user-defined functions (UDFs) to extend SQL and the lack of support for in-process or out-of-process analytic processing. Nor is there any native connectivity to geospatial and unstructured data – which is only achieved through custom connectors either built by customers or sourced via third-party technologies. Additionally there is no support for MapReduce programmatic access to data. We also find that the proprietary way in which Infobright compresses data for queries does limit its capability to integrate with a broad third-party BI front-end tooling, which is optimized for star schemas. That said, Infobright does integrate with open source BI platforms like Pentaho, Jaspersoft, and Actuate.

Practical scaling limitations – A general weakness for Infobright has been its single-server-centric, no-MPP architecture, which means that a query is tied to one host CPU server only, with no logical partitioning across multiple servers, which hinders SMP/MPP scale-up. However, Infobright is starting to address this in the latest versions; IEE does come with data replication, based on MySQL statement-based replication, which provides a degree of scale-out by spreading the load among multiple servers to improve concurrency and performance. The new Infoplance configuration does offer scale-out/up configurations – from 12 to 144TB of data in a single appliance node – with no additional hardware requirements. Infoplance also supports user scalability by adding additional nodes to the Infoplance cluster, as well as high-availability deployments.

Evolving database management tools – Infobright's database management tools – which it inherited largely from its MySQL heritage – generally lack the breadth and sophistication of rival analytic database platforms. While MySQL (and its partners) continue to make advances in database management tooling, around areas like availability replication, workload management, security, and database monitoring, MySQL is still designed as a quickly and easily embeddable database. Better capabilities around query engine diagnostics (including automatic query throttling), database fault tolerance, failover, load balancing, and high-availability capabilities, which albeit have not typically been a focus in machine-generated data use cases, need to be better supported, especially as Infobright's expanding customer base will bring these issues to the fore. Since many third-party software vendors are reluctant to commit to working with small software companies, this puts even more pressure on Infobright to develop its own management software capabilities. There is nothing inherent in Infobright's architecture to preclude customers from adding their own failover, high availability, etc. with off-the-shelf components. However, this will require custom development. While the newly introduced Infoplance potentially addresses many of these concerns, it will take time to validate its effectiveness. The planned support for Postgres promises to deliver more sophisticated enterprise-class database management capabilities, which Infobright can leverage in its future development.

Narrow focus on machine data and possibly limited mainstream market appeal – Although technically similar to how other columnar databases approach data management, Infobright's architecture is focused primarily on machine-generated data. As such, Infobright maintains a low profile against rivals, such as Vertica (now part of HP), Greenplum (now part of EMC), and ParAccel (now part of Actian), who have a broader focus beyond machine-generated data and are positioned as general-purpose analytic systems. Customers looking for a more generic analytic database system that works well against volatile data to support applications around customer or inventory analytics that require constant in-place updates and appending of individual data records might feel restricted by the Infobright engine, which is designed to quickly compress, analyze, and load large amounts of data that will remain static.

Current dependence on MySQL – Infobright is currently engineered to work on the MySQL database. As such, the company has a vested interest in the continued development of MySQL, which is owned, of course, by Oracle – a potential rival in the analytic database space. While Oracle continues to enhance MySQL, the expiration of Oracle's agreement with the European Union (EU) is drawing closer, and with it the possibility that Oracle will seek to change agreements with OEMs. This might present a long-term risk for Infobright's ability to advance the product's functionality and its customer base. Recognizing this, Infobright is planning to soon release a Postgres version, which will provide a truer

open source strategy. The company says that customers across both platforms will experience similar functionality.

Opportunities

Partnerships – The simplicity afforded by Infobright's column-store design and Knowledge Grid metadata layer – which reduces the dependence on special index and aggregate creation and specialist tuning activities – continues to attract numerous OEM partnerships (which remain a big part of Infobright's business), for BI and data warehousing and potentially as a base for a vertically specialized database platform. To date, Infobright has signed up some 30 partnerships – including open source providers like Pentaho, JasperSoft, and Talend, as well as non-OEMs like Polystar, Mavenir, and JDSU – to embed a database that requires little support on their part and which also adds meaningful credibility to the solution in certain business contexts. The company is also gaining traction as a DBMS OEM for software vendors that analyze machine-generated data. The planned support for Postgres will help, as Oracle's MySQL strategy is not particularly well set up (from a licensing perspective at least) for an OEM model for embedding MySQL into third-party appliances and cloud infrastructures. It is noticeable that Infobright is not currently partnering for the sales and support of the appliance, which will limit its early visibility and the ability to provide support for hardware issues in early deployments.

Anchoring an open source analytic reference architecture – Infobright's open source strategy and roots certainly hold great appeal to organizations looking to build a high-performance analytic ecosystem system – akin to a LAMP-like stack – around MySQL. Storage engines are arguably the weak point of MySQL, which Infobright addresses directly with its solution. But while the company's use cases are a good match for open source, it is not the only option. Rival, albeit more general-purpose, solutions are being built around MySQL (Kickfire), MonetDB, and most notably Postgres; vendor solutions such as EMC Pivotal, IBM Netezza, Microsoft DATAlegro, Vertica, ParAccel, and Teradata Aster Data are all based in one way or another on PostgreSQL (even though Vertica includes no PostgreSQL code). However, to varying degrees, they have inherent limitations for storing and analyzing machine-generated data.

Threats

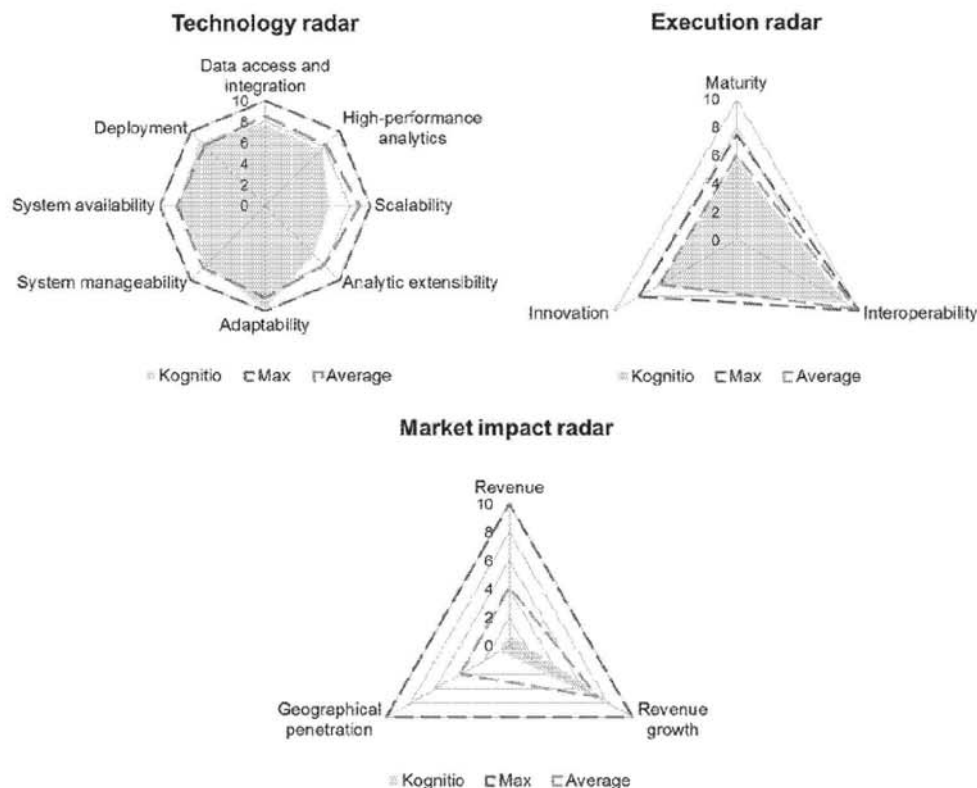
Competitive standing in the market – Infobright is still a relatively small player among its analytic database peers and has been slow to achieve revenue growth; it is privately held, with roughly 50 employees (in Canada, the US, and Europe, with a development team in Poland). Investors – including RBC Ventures and Flybridge Capital Partners – injected capital into Infobright in 2008 and the company relies on additional funding rounds to build out its infrastructure and capabilities. That said, Infobright has a growing customer base of over 484 direct and OEM IEE and Infopliance clients, and its ICE offering boasts over 100,000 downloads – including over 85,000 registered community participants. However, the company still lacks a sizeable corps of vertical application and service provider partners. To drive growth Infobright needs to show that its open source model is not just about cost reduction, but that it is good enough for organizations to license the IEE commercial version, in order to drive revenue.

Risk of open source model becoming less differentiated in the market – Infobright is constantly being challenged to differentiate its offerings and open source model from mature column-store analytic databases that offer column-store and column-compression capabilities and alternatives. In particular,

as analytic workloads become more mixed and complex, Infobright must act quickly to manage these workloads – something that it does not do particularly well today. Alternatively, Infobright's distinct technology – particularly its patented Knowledge Grid and strong column-store compression technology – might also make it a likely acquisition target by a larger columnar database vendor.

Kognitio (Ovum recommendation: Challenger)

Figure 9: Kognitio radar diagrams



Source: Ovum

Ovum SWOT assessment

Strengths

Rapid analytic processing – Low latency and high throughput result from Kognitio's in-memory scale-up/out architecture, and are particularly effective for iterative ad-hoc and complex queries. Kognitio's "Massively Parallel In-memory Code Execution" engine allows analytical algorithms (written in any language) to be embedded within standard SQL wrappers and executed in a fully parallel context. Kognitio has delivered multiple architectural options for getting data from a range of sources on and off disk in the simplest possible way, bypassing the need for the overheads of indexes or materialized tables, and thus bringing more horsepower to bear against scaling data volumes. This

gives users more freedom for spontaneous data exploration and analysis because they are not confined to analyzing data in a preordained way.

Cloud provisioning – Kognitio's extensive experience in cloud deployment is apparent in its marketing – the aim being to make high-performance analytics more affordable to companies of all sizes. The Kognitio Cloud, which is attracting most new business, can implement scale-out multi-node instances of Kognitio platform on both private and public (AWS) cloud infrastructures. The latter uses AWS' subscription-based expense cost model and on-demand usage model, and also taps into newly introduced EC2 High Memory Instances with 244GB of RAM, which Kognitio claims offers significantly lower cost.

In-memory data loading and management – Kognitio Analytical Platform has a mature data access strategy that is not just about caching data in memory. It supports both bulk and streaming data ingestion, configured easily via a graphical interface, that pins data into RAM close to the CPUs in low-overhead structures, where the dynamic random access nature of RAM can be exploited by all CPU cores participating in query and analytical processing. Tables can be created purely RAM-only, or (optionally) be synched back to disk or external stores for persistence. The RAM image of a table may only hold a subset of data with the optimizer able to intelligently direct-query access based on the predicates – this allows large fact tables to be stored on disk with only the recent "hot" (i.e. actively accessed) data pushed and analyzed in-memory as read-only snapshots of the underlying query data. All data views are optimized in RAM, with hashing, partitioning, sorting, replication, pre-joins, and compression for performance, but can be changed according to application use cases. Users can choose to leverage the optional disk subsystem Kognitio offers for free (product license costs are only calculated on the amount of RAM used). When doing so, the system can write data both to memory and disk, then keep them in synch.

Flexible pricing model – Kognitio gives considerable flexibility to customers to switch painlessly and economically between the various physical deployments of the system – on-premise and cloud. Rather than use a traditional price-per-terabyte model, Kognitio only charges for the available memory (RAM). Specifically, it has pinned AWS Marketplace software license costs to one cent per gigabyte per hour. Kognitio also offers a "free" version for small deployments with perpetual use licenses for up to 128GB of data in RAM (persistent storage is free and unlimited).

Weaknesses

Data compression remains a concern – Ovum has some reservations about Kognitio's in-memory compression strategy because the platform is a row-based database. While Kognitio has long argued that the CPU cycle cost of compression is not justified by the RAM savings from compression, it has finally relented and only recently introduced in-memory compression in the latest version of the product, which has bumped up scalability. The company claims its system has been configured for clients at sizes of up to 64TB of RAM (128 nodes) and that linear scalability has been proven in testing. However, it notes that clients who want the fastest performance possible may choose to disable this feature, leading us to believe that its compression strategy is still immature and could put Kognitio at somewhat of a disadvantage versus columnar products when dealing with data sets that are only going to grow.

Lacks native packaged in-database analytics functions – Kognitio Analytical Platform would benefit from a richer set of built-in modeling or analytic functions. It does support complex SQL analytics (ANSI 2011 standards) along with its massively parallel in-memory code execution capability to run any language embedded within SQL. While this offers a lot of freedom for end users, and allows direct exploitation from tradition BI tools, it is not yet packaged into ready-to-use, defined use-case solutions for business users. However, predictive functions like R are more complex; Kognitio therefore needs to invest in partner-defined modules or algorithms and/or build native analytic function libraries into the platform so that users can more easily exploit these advanced algorithms.

Integration with third-party tools – Kognitio does not offer any front-end BI and analytic tooling, relying exclusively on custom or partner-led integrations using ODBC, MDX, ODATA, and XML/A. While some of these integrations have been "certified," interoperability could be updated (most certifications are for older versions), tightened up, and made easier. Additionally, Kognitio lacks a broad base of systems integrator and consulting partners that have vested skills in its platform – an issue not uncommon for a small vendor with relatively low visibility in the market. That is also reflected by a distinct lack of process- or vertical-specific analytic solutions available today.

Lack of columnar option – Kognitio is architected – as an administrator-led function – to organize data in a row-based scheme in memory. While this has benefits in terms of ease of updates, it does have a relative disadvantage to columnar layouts when data is inserted into tables or read. Columnar certainly benefits analytics with quicker and more flexible read strategies and physical storage compression – something that rival in-memory analytic databases are already embracing with hybrid storage approaches. Arguably, many types of business analytics prefer to work at an aggregate level, calling into question the need for granular row-based data access and analysis – in-memory or not. Recognizing this, Kognitio has developed a token option for holding small tables in a columnar format.

Long-term growth – Kognitio remains challenged by a relatively low profile and visibility in an increasingly competitive space. Impressive growth figures – 99% from fiscal year 2013 to 2014 – are reflective of a modest Europe-centric base of around 50 (albeit fiercely loyal) customers. Overall, growth has remained sluggish; the company added 12 new clients in 2013, including 8 in the US where Kognitio continues to struggle against larger rivals with bigger local marketing presence and resources. While Kognitio has been in business for 25 years, the company remains challenged to grow its business in these key markets. In particular, one area that Kognitio has not tapped effectively is partnerships, particularly with US-based service providers.

Opportunities

Extending the value of existing data warehouses – The greatest short-term opportunity for Kognitio is to reinforce its position as a specialist in-memory analytic platform to complement existing EDW and Hadoop deployments. It should aim to target a smaller subset of business analytics that would benefit from accelerated and cost-effective in-memory analytics that would otherwise eat up valuable compute resources.

Leverage cloud to expand SMB business – Historically, the lion's share of Kognitio's business came from software or appliances. That has now changed – today over 70% of its pipeline is cloud-based, focused on its public and private cloud implementations. A continued shift towards cloud-based analytic

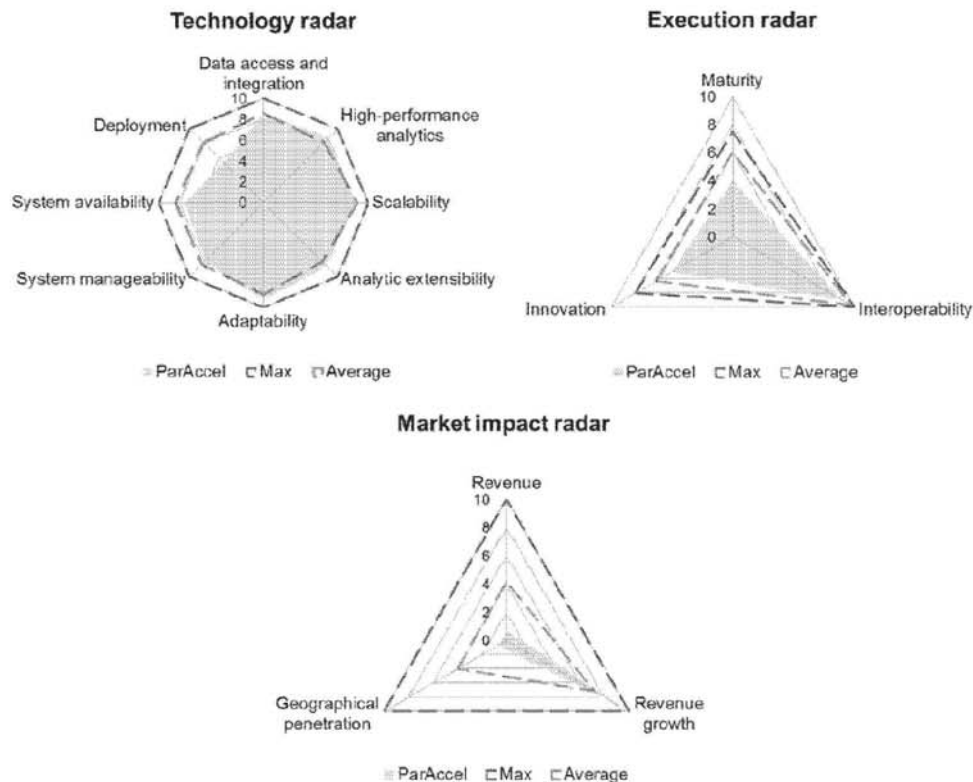
strategies by the highly price-conscious SMB segment should continue to open up further opportunities for Kognitio to deliver attractive price-performance analytics on server farms in the cloud.

Evolving Hadoop strategy – Kognitio has been slow to devise an integration strategy for Hadoop. However, the latest version of its platform now comes with direct HDFS connectivity features that allow Hadoop data to be transported in massively parallel transfers directly into in-memory on Kognitio processing nodes, as well as pushing portions of query processing down to the Hadoop nodes. This should open up new opportunities in Big Data environments, particularly in organizations that want data-processing options driven by standard SQL. By sitting on top of the HDFS and interacting directly as an in-memory processing platform that can interact in SQL, Kognitio affords the same capabilities as "SQL on Hadoop" applications with the added ACID-like features typically associated with a data warehouse. Kognitio has also developed some novel Hadoop integration, notably the ability to push down basic filtering and projection into HDFS and then bring back into memory only the rows and columns required for a query, which can often number into billions of rows. Finally, Kognitio's recently announced capability to parallelize R scripting could give Kognitio further differentiation in the Hadoop predictive analytics marketplace.

Threats

Competition is heating up – Kognitio faces a rising tide of competition from smaller analytic database specialists as well as large IT vendors offering rival (but different) in-memory and "hybrid" columnar architectures. Today, Kognitio is strictly a row-based in-memory database, and should invest more R&D into extending its legacy row-based-only architecture; notably, the company has plans to add hybrid columnar functions in the future. Nor is Kognitio the only analytic database solution targeting the SMB space; it faces stiff competition from the likes of Amazon Web Service Redshift (which integrates a solution with ParAccel – a competitor in its own right) and others.

Narrow ecosystem and stack – Kognitio lacks the broader ecosystem and software stacks of rival analytic database providers. As such, the solution might well be seen not as a standalone analytic platform, but a specialized component of a broader best-of-breed analytical environment. In this respect, vendors like Oracle and SAP hold an advantage because they can effectively heavily discount, or in some circumstances even give away, part of their software stacks in deals. Kognitio does not have a portfolio extensive enough to afford its customers that luxury.

ParAccel (Ovum recommendation: Challenger)**Figure 10: ParAccel radar diagrams**

Source: Ovum

Ovum SWOT assessment**Strengths**

Query optimization engine – The trump card of ParAccel's MPPdb database is its patented query optimization techniques – built on some clever reengineering of the open source Postgres SQL to compile rather than interpret queries. Its Omne Optimizer engine ensures uniformity of processing speed and response across a wide range of analytic queries. Version 4.0 includes several refinements around aggregation, data interconnects, I/O optimization, data load prediction, and efficient in-memory allocation that further boost query performance. Workloads that ParAccel MPP excels in include complex SQL queries that need to process data across large data domains as well as ad hoc and discovery-like applications. But, in line with its narrow focus on speed, ParAccel MPP only offers optimization for in-process analytics execution.

Extensible library of analytic functions – ParAccel MPPdb offers an extensive library of more than 500 pre-built functions that run directly inside the database. Leveraging partnerships with Fuzzy Logix

and Numerix, the company has also authored its own proprietary library of analytic functions that ship with the core platform (many of which are aligned to core vertical markets, such as financial services and digital media) and which also help accelerate development of customized analytics models. What's currently missing, however, is full parallelization across these functions – an issue that Actian is addressing. ParAccel also offers APIs to extend analytic functionality with an extensible framework for defining user-defined analytic functions (UDFs), into which all analytic processes execution capabilities are subsumed.

Concurrent query scale-out – The database handles high-volume analytics well for highly concurrent, multilevel, highly recursive query environments. Actian references clients that scale their databases from the 20–50TB range and claims its solution can easily support up to 5,000 simultaneous clients with concurrent connections, making it suitable for sizable enterprises. For example, one of its retail customers has an algorithm running within MPPdb with 10,000 lines of code. Another financial services customer runs a 25,000-line SQL query for dynamic risk stress-testing.

SOA-based data plumbing – MPPdb's bidirectional ability to easily and quickly move analytic data and processes in and out of data warehouses and Hadoop is a key strength. The company offers a range of On Demand Integration (ODI) high-performance connectors that effectively treat these sources as engines with its Omne Optimizer query optimizer as a front end. In particular, MPPdb also supports the "WITH" construct in SQL syntax (a useful capability that some of its competitors still lack) and Apache HCatalog, the emerging metadata facility for the Hadoop stack which allows users to perform analysis on data warehouse and Hadoop data in a single query.

Weaknesses

Potential data-loading issues – While data can easily be moved in and out of the database, what had been neglected are processes and interfaces used to load data into the analytic database store in the first place – which is significant for a columnar-based product. Actian has already addressed this issue by integrating MPPdb with its Dataflow engine (formerly Pervasive Software's DataRush), which allows companies to prepare and enrich data in one flow within an orchestrator that accesses all data sources, then loads data into the MPPdb system. Ovum advises customers to test data loading while conducting a proof of concept for this database and thereafter make their decision.

Lightweight platform management – We find ParAccel's logical data model creation, metadata management and storage, and mixed workload management quite rudimentary compared to rival columnar databases. Actian actively promotes that this is a benefit of the unique ParAccel MPPdb architecture; its optimizer accomplishes a hefty amount of work in execution, which mitigates the need for heavy database management and also requires significantly less database administration resources than other analytic platforms. The company is starting to focus more sharply on partnerships to achieve this end. Ovum believes that it can also provide these capabilities through externally managed services – such as in the AWS cloud.

Slow market traction – ParAccel MPPdb has experienced relatively slow growth over the past several years. Even after receiving tens of millions in venture backing (including a sizable investment from Amazon), its market presence remained static in the past. However, it has been reporting exponential growth in its pipeline and market presence following the Actian acquisition. Since ramping up its sales

organization and entering into the Actian portfolio, ParAccel MPPdb has garnered close to 100 customers – the overwhelming majority of which are in North America. Ovum believes that the company must accelerate its revenue and numbers of new customers more quickly than it has so far to ensure long-term stability and viability. The acquisition by Actian has undoubtedly raised ParAccel's visibility in the market. But Actian has its work cut out to rationalize and focus its resources across the product segments it is pursuing (Big Data analytics, cloud integration, and core database management). Until that is resolved, Ovum retains a cautiously optimistic view of ParAccel MPPdb.

In-memory option still evolving – While ParAccel MPPdb allows users to persist all data and processing in memory, the platform is architected to be disk-based in nature. However, Actian reports that one of its MPPdb customers is running more than 60TB in memory. The company offers an option to pin certain data tables into its memory cache for rapid processing, giving users the option for in-memory and disk approaches. Ovum finds that the in-memory option in ParAccel MPPdb is still evolving and potentially risks complicating the setup of a system. For example, data cannot be persisted in-memory throughout an entire analytic processing task, and if users want a data structure to persist past the end of a query in memory, it needs to be physically mapped to a row architecture. ParAccel is still working on ways to make this more natural and less prone to governance abuse. In future releases it promises more flexible UDFs in terms of what kinds of data structures can be managed and analyzed in memory.

Opportunities

Establish a deeper foothold in Big Data analytics – Actian gives ParAccel a firmer foothold in the Big Data market, albeit as part of a product line spanning database, integration, analytics, and applications. That said, ParAccel's technology is well positioned to ride the Big Data market wave, and the company is increasingly positioning itself as a platform that brings together the worlds of SQL and Hadoop analytics – witnessed by recent efforts to extend its ODI capabilities to support Hadoop, leveraging HCatalog and HDFS. However, ParAccel is not the only vendor to develop this capability. Eventually Hadoop integration will become commoditized, and the risk of not advancing and innovating quickly in this area carries the risk of being overshadowed by rivals that are now part of much larger IT vendors.

Recently, Actian undertook a few initiatives which demonstrate its commitment to Big Data and Hadoop. In October 2013, Actian and Hortonworks jointly announced an integration between the ParAccel Big Data Analytics Platform and the Hortonworks Data Platform to address adoption hurdles of data extraction and poor run-time performance.

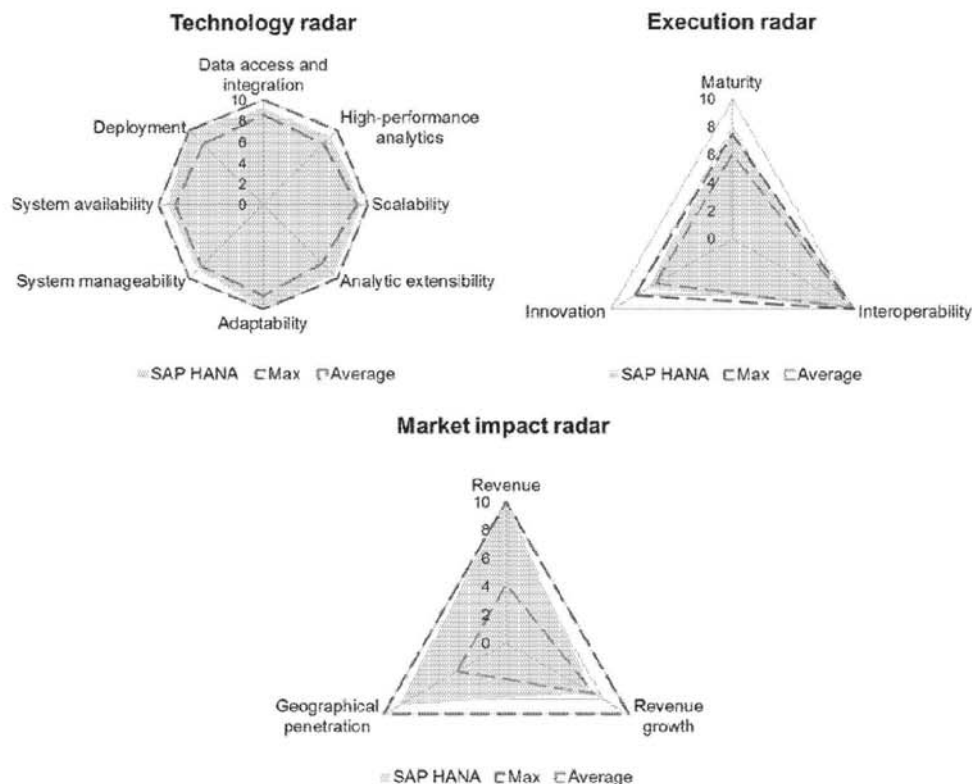
Leverage partnerships to expand market reach – Establishing a good set of partners is an important part of Actian's strategy. The company's investment in AWS, and licensing of components of its analytic database for use in its RedShift data warehouse as a cloud service, gives Actian a potentially lucrative OEM licensing revenue stream and lends stronger credibility to customers concerned about dealing with a small company. That, coupled with a relationship with MicroStrategy (which bundles the ParAccel database as part of its own cloud portfolio); deepened partnerships with Dell, NetApp, and Cisco; and continued on-boarding of resellers and service providers (over 40 were added in 2012) should lay a solid foundation for seeding a critical mass in the market. However, Actian needs to create momentum from these initiatives; many columnar database rivals boast more OEMs (and more customers) than Actian does today.

Make the whole more than the sum of the parts – The MPP technology will benefit from being part of a larger Actian business that already focuses on analytics as part of its product strategy. Technically, Actian has an opportunity to fuse the former ParAccel and Vectorwise technologies to produce a lightning-quick and scalable (conceivably petabyte-size) analytic platform that melds Vectorwise's on-chip, SMID process optimization with ParAccel's scale-out capabilities. Both technologies could be complementary and provide missing pieces that the other lacks; for example, MPPdb seems to struggle with single-node use cases while Vectorwise is better at them (until queries get more complex). MPPdb also has the potential to offer greater scale, more analytic platform capabilities, and mature data-plumbing capabilities. Actian also has over 14,000 customers on a product line spanning databases, data warehouses, integration, and applications – a base that offers considerable opportunities for cross- and upselling of the MPPdb platform. That, however, will require sales synergy between the composite products under Actian's wing.

Threats

Grasping opportunities in the cloud quick enough – ParAccel MPP's cloud offerings have continued to evolve; its maturity is evident from commitments made by leading analytic tool vendors like AWS, MicroStrategy, and Birst. Since the acquisition by Actian, its cloud partnership strategy has expanded. The acquisition and evolving partnerships undoubtedly offer considerable opportunity but we see several caveats: AWS, to our understanding at least, only offers a stripped-down version of the MPP engine that does not carry over MPP's proprietary optimization profile that well (high-speed connectors for Oracle and Teradata databases and Hadoop have not been licensed by Amazon), hence making it hard for RedShift customers to switch to it from any other relational DBMS. Redshift customers looking for additional functionality must purchase the full ParAccel MPPdb product from Actian. Ovum believes that ParAccel must take an aggressive stand on the cloud over the next year if it is to expand its market share.

Market perception under Actian brand – ParAccel is the newest part of Actian's goal of assembling a comprehensive database and analytics stack that includes a data integration engine and different data platforms, also comprised of technologies acquired from Ingres, Pervasive, and Versant. Ovum believes that ParAccel is likely to supersede Actian's Vectorwise as its go-to-market analytic database. Building a database conglomerate is tough, and indeed has proven futile in the past (e.g. Informix). While Actian is starting to raise awareness of its analytic portfolio expansions, it has a poor track record and is not great in this respect; it has not accomplished much with Ingres (its original acquisition) or invested enough in Vectorwise to make it a major market player. The question is whether Actian is likely to start now that it owns ParAccel as well. Integrating Vectorwise's strengths into the ParAccel engine will not be straightforward. Right now, Actian has not even suggested such a move, but it needs to act quickly to clarify the roadmap and positioning for both of its analytic databases.

SAP HANA (Ovum recommendation: Leader)**Figure 11: SAP HANA radar diagrams**

Source: Ovum

Ovum SWOT assessment**Strengths**

Faster processing because parallelism is engineered at the (Intel Xeon) chip level – Parallelism in HANA extends from the internal Intel Xeon instruction set to the multi-node scale-out clusters which can handle very large volumes of data. For business users, this translates into faster data loading and processing even when working with very large data sets. Intel and SAP have invested significant engineering effort to optimize SAP HANA for Intel Xeon processors, ensuring that operations in HANA have the advantage of the latest innovations in the processor. For example, HANA can use SIMD instructions (chip level) to perform an operation on several values in parallel. In addition, SAP HANA's allocation of separate threads to individual columns for dynamic aggregation and materialization helps it use all the cores on a single machine for a query if needed. This concept extends across nodes in a cluster where data is automatically partitioned and queried in an optimized, distributed manner to use compute power from an in-memory cluster of commodity hardware.

"Powered by HANA" is a new direction of SAP's strategy – Apart from BI/analytics, SAP HANA is now a key ingredient in many of the applications and platforms in SAP's vast portfolio, such as SAP Business Suite (packaged business applications containing ERP, CRM, SCM, SRM, and PLM), parts of EPM, GRC, and visualization (Lumira). As a result, SAP HANA's strengths are not only as a standalone analytical database but also as an important enabler for SAP's upcoming product roadmap.

Ovum believes that HANA is productive when acting as the core platform/database running SAP's applications, primarily due to ease-of-integration and performance enhancements that SAP HANA brings to each application area. Business functions that show immediate benefits from SAP HANA adoption include fraud management (part of GRC), business planning, and optimization (part of EPM).

Partnerships for implementation, support, and customization are strong – SAP has developed a comprehensive set of services, rapid-deployment solutions, and trained implementation consultants for HANA from its large ecosystem of partners. Its rapid-deployment solution is designed for customers to go live in less than six months and offers preconfigured software, implementation services, content, and end-user enablement for a fixed price and scope. Its solution accelerators, on the other hand, are targeted at specific functions such as marketing (Customer Segmentation Accelerator) and finance (Finance and Controlling Accelerator) and offer quick deployment for specific use cases. Ovum's research also reveals that major system integrators and IT services vendors have devoted significant resources to developing HANA talent, which will ensure that SAP HANA skills are abundant in the market so customers can deploy, manage, and tweak the solution to their needs.

SAP HANA and mobile analytics go hand in hand – Packaged and consumer-like mobile reporting/analytics on top of HANA (both on-premise and in the cloud) is a growing use case for the database. An in-memory database layer provides an ideal data-processing back end for mobile platforms (such as Sybase Unwired Platform). It is particularly useful when querying and calculating across medium to large data sets, providing subsecond responses that mobile application users are accustomed to. However, while in-memory allows data to have a smaller footprint, SAP HANA in-memory is always going to be more expensive than databases that work on disks. It is important that customers work out the cost/benefit for SAP HANA for their mobile use cases.

In-memory systems such as HANA will be especially useful as the functionality of mobile applications becomes more complex. Most of today's mobile device hardware and mobile clients are ill-equipped to handle data exploration, advanced analytics, and predictive capabilities across large subsets of data. HANA feasibly opens up a new set of sophisticated analytic functionality that is accessible from these devices.

Weaknesses

Predictive analytics offerings still to attain maturity – SAP has traditionally engaged IBM as an OEM partner for predictive analytics, releasing its own predictive solution (SAP Predictive Analysis) only in 2012. The software is built on a combination of in-house development and open source R components, and taps the SAP HANA database for fast, in-memory analytics processing. While the growing popularity of R in predictive analytics is indisputable, Ovum believes that SAP will need to put in continued efforts to optimize R algorithms to best exploit its in-memory architecture. Ovum would also like to see further expansion in SAP HANA's internal predictive libraries to include more

industry-specific predictive metrics and analysis. At present, SAP HANA has a native in-database Predictive Analysis Library (PAL) which can support 24 native predictive functions and a Business Function Library (BFL) that can support 60 native business functions.

SAP announced its intention to acquire KXEN, an independent predictive analytics vendor, in September 2013. Ovum believes that while this bodes well for SAP's predictive capabilities, it will be at least a year before the vendors can integrate their technology.

Needs better unstructured and semi-structured data analytics capabilities – SAP HANA would benefit from strengthening its unstructured data analytics capabilities, especially around text analytics. A majority of unstructured or semi-structured data that is available and analyzed today comes from text feeds (social media, clickstreams) and enterprise content management or enterprise search management tools. To analyze such data effectively, SAP must develop strong text-processing (including natural language processing) and mining tools that operate on top of SAP HANA and use its in-memory and parallelization capabilities.

The vendor is already making progress in this direction. Since SAP HANA SP05, search and text analysis capabilities include document filtering, which unlocks textual data from binary formats such as Adobe PDF, MS Office, Apple iWork, and Lotus Symphony. The vendor released SAP HANA Text Analysis in 1Q13. Ovum would like to see further development in this module to include extensive dictionaries, rules, and customization for industries. The vendor will also need to provide pre-built models/content that enable statistical approaches to text mining.

Ovum also finds SAP HANA's capabilities slightly underused when dealing with unstructured or semi-structured content that resides inside a company's firewall – emails, documents, videos, etc. While SAP has the capabilities to harness these sources, it needs to better integrate these capabilities into SAP HANA and offer targeted solutions. Currently, SAP provides a UI for the Information Access toolkit that ships with SAP HANA and allows customers to combine analytics, search, data mining, text mining, spatial, and graph into application front ends.

For effective mining of in-house semi-structured data on SAP HANA, Ovum believes that SAP could also prioritize integration with a strong enterprise content management (ECM) solution and an enterprise search solution to work directly in-database with SAP HANA.

SAP has traditionally not been an active player in both the ECM and enterprise search markets, tying up with Open Text for most of its ECM needs. This is an area ripe with opportunity, which Ovum expects most analytic database providers to pursue aggressively.

SAP's cloud offerings need a few more tweaks – SAP recently released its SAP HANA Enterprise Cloud offering, a managed private cloud service that helps customers migrate on-premise SAP projects to the cloud. To access this cloud, customers need to possess a full HANA enterprise license. Ovum believes that this licensing strategy has both positives and negatives. On the one hand, it provides customers the convenience of a private cloud to use for quickly demonstrating PoCs (proofs of concept) to running production environments, while SAP takes care of provisioning and infrastructure. On the other hand, the requirement of possessing a full HANA license to operate on the cloud implies that customers cannot scale down their license (but can still scale down their subscription of infrastructure) in response to changing user and business requirements. SAP specifies that its other cloud offerings, such as SAP HANA Cloud Platform, SuccessFactors, Ariba, Cloud for Travel, and Cloud for Financials

are designed to cater to demand elasticity, while SAP HANA Cloud is targeted at simplifying SAP HANA deployment and management for larger installations with a long-term perspective. In other words, SAP HANA Cloud is meant for organizations that consider SAP HANA to be an integral or mission-critical part of their analytic processes and will not consider a scale-down in the near future, but could use some help in maintaining and rolling out the solution.

Opportunities

Target native user base and displace third-party vendors from the database/analytics stack –

With SAP HANA, SAP finally has a viable option to prevent other vendors from infringing on its ERP stack. Traditionally, customers purchasing SAP applications (Business Suite, ERP, CRM, etc.) deployed third-party database vendors such as Oracle, IBM, and Microsoft. Given their proximity to the data collection and storage processes, these vendors influenced BI or data warehouse purchases. SAP should aggressively target organizations with such mixed technology stacks, quoting the performance improvements (realtime advantage) and simplicity of running SAP Business Suite on HANA (the benefits include native integration, ETL performance improvements, and subsecond querying). Ovum notes that these benefits only accrue to organizations that have the new versions of Business Suite optimized to exploit in-memory architectures. Ovum estimates this market – organizations with mixed technology stacks – to be a huge opportunity that could well lead to sustained double-digit growth for SAP HANA. To sweeten the deal, SAP should also continue to develop application modules that work only with SAP HANA (such as Fraud Management, a part of SAP GRC).

Develop industry- or function-specific offerings (with partners) – SAP's current approach is to focus on developing SAP HANA as a horizontal solution. However, Ovum believes that future growth opportunities for SAP will lie in creating specialized pre-packaged appliances that target either vertical industry- or process-specific analytic needs. Such bundles would ideally comprise a set of pre-built industry data models, data integrations into specific sources, and pre-built visualizations/dashboards specific to the industry and/or function. Currently, the vendor has only developed a few accelerators and "powered by HANA" modules to that end.

Threats

Lower-than-anticipated uptake for realtime applications – SAP is investing heavily in SAP HANA and its in-memory technology. One of the key value propositions of SAP HANA is its speed, which is of great value to realtime applications. However, many enterprise application areas might not benefit significantly from realtime operations/analytics (for example, the accounting department for monthly salary disbursements). Also, many enterprises may not have processes that can exploit the ability to use speed to add new flexibility that actually results in business change. An abundance of such enterprises in a slow economy could lead to elongated sales cycles for SAP HANA.

Ovum believes that a further slowdown in the economy could result in organizations delaying their upgrade cycles, which might also impact SAP HANA's projected growth slightly more than its competitors given that SAP HANA typically requires higher upfront investment.

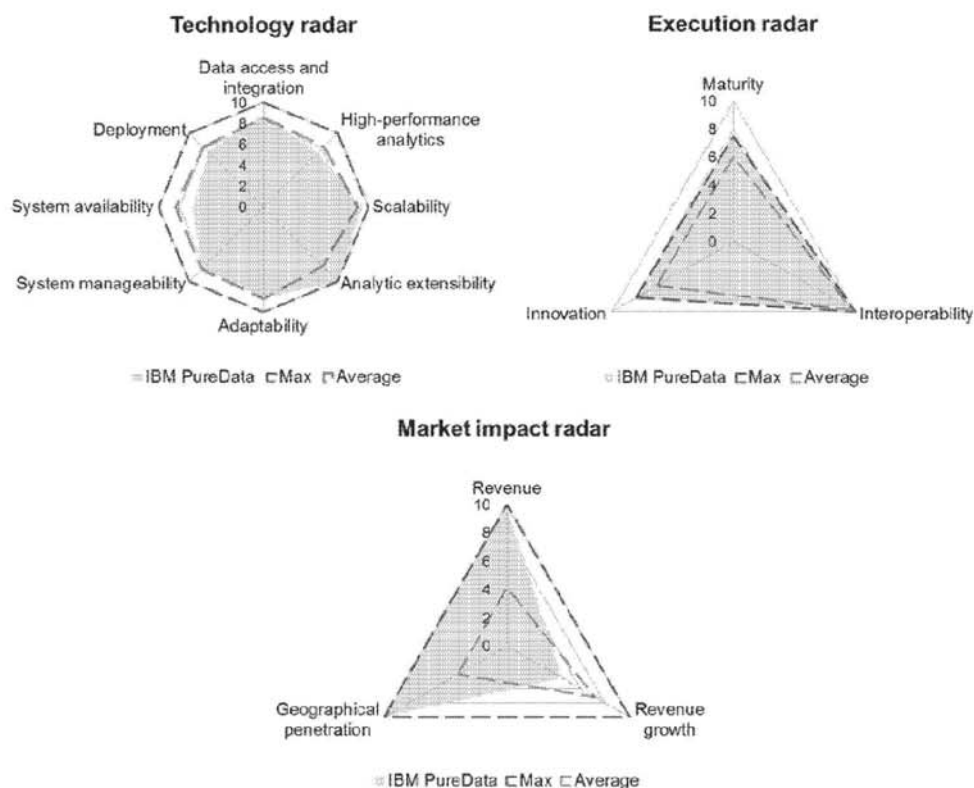
Immensely competitive and confusing market – The analytics market has been shaken up by the emergence of both Big and Fast Data stores. Realtime and Big Data analytics can be served by many

different types of platforms, including in-memory, massively parallel processing (MPP), columnar, and NoSQL databases, each with its unique advantages, drawbacks, and best use cases. On the analytical side HANA appears best suited for structured data analytics; for OLTP, it is best for applications that are optimized for it (although it is compatible with most apps that store data in SQL). However, for most analytical use cases HANA faces very stiff competition from MPP and columnar databases. The high growth potential of the market has resulted in a vendor frenzy, with almost all major enterprise software vendors establishing a beachhead and claiming that their approach is the panacea to all analytic problems.

Given the nascence of the segment and the lack of unified benchmarks (such as TPC-H), it has also become increasingly difficult for end users to judge the merit of individual solutions and to make viable choices. Ovum believes that in this phase users are overwhelmed by the variety of options for storing data and analyzing it. SAP needs to continue to invest significantly in communicating the benefits of SAP HANA while ensuring that it demonstrates business value over technological prowess.

IBM PureData System for Analytics (Ovum recommendation: Leader)

Figure 12: IBM PureData radar diagrams



Source: Ovum

Ovum SWOT assessment

Strengths

Mixed-workload performance – Performance advantages are gained from Netezza's field programmable gate arrays (FPGAs) that uniquely combine IBM blade servers and disk storage with IBM's patented data filtering, which eliminates the need for defining and optimizing indexes. Thanks to a recent refresh of the underlying Netezza NPS/TwinFin architecture – which takes advantage of the new N2001 Striper architecture – users benefit from significant improvement in query performance, notably a 3x improvement in scan rates for a 128Gbps effective scan rate, and (significantly) much better mixed workload management and concurrency for simultaneous queries. Notably, IBM has also leveraged the Netezza core technology to create hardware accelerators for other IBM platforms; modified versions are also available to offload complex analytical queries from DB2 mixed workloads onto IBM System z mainframes using the z/OS version of DB2.

Range of in-database analytic functions – PureData System for Analytics excels in providing advanced algorithmic processing functionality inside the database – which allows data modelers and analysts to operate on the data directly inside the appliance instead of having to offload it to a separate infrastructure and deal with the associated data preprocessing, transformation, and movement. The system ships with the largest library of analytic functions in the market, including over 200 built-in mathematical, statistical, transformation, time-series, data mining, predictive (PMML 4.0 model support), scoring, and geospatial functions that execute analytics in parallel. The library is extensible through third-party options from SAS (for scoring), Fuzzy Logix, Revolution Analytics (for R-based analytics), Zementis (for PMML), and ESRI (for geospatial analysis), as well as SDK for custom creation (using a number of different programming languages, including C/C++, Java, Perl, Python, R, and Fortran). Finally, there is also a SQL Extension Toolkit for PureData System for Analytics, available from IBM's Passport Advantage and Fix Central websites – available free of charge to customers. Analytic function extensibility is also supported via Eclipse plug-ins, UDFs, and MapReduce integration.

Rapid deployment through "integration by design" – The system integrates advanced analytical capabilities in a single, easy-to-deploy system – in hours, not days – and with minimal upfront design, integration, and fine-tuning. IBM touts PureData System for Analytics as an "expert-integrated" system that simplifies how it is procured and deployed – meaning that it is delivered as an optimally configured, assembled, and integrated system for running advanced queries. It also provides standard interfaces – through standard ODBC, JDBC, and OLE DB interfaces – to best-of-breed BI and data integration tools and offers connectivity to Big Data platform components like Hadoop. Integrations with IBM Cognos, IBM SPSS, and InfoSphere Information Server are, not surprisingly, kept very tight as a design point.

Scalability across heterogeneous environments – PureData System for Analytics emphasizes enterprise data warehouse scale and/or deployment against multiple database platforms. Because IBM is a relatively platform-agnostic vendor, customers with multiple database platforms, such as IBM DB2, Microsoft SQL Server, and Oracle Database, will find PureData System for Analytics a better fit for heterogeneous database environments. The system also comes in various sizes, from entry-level configurations that can incrementally be scaled up on IBM servers running industry-standard x86 Intel chips. However, its asymmetric and parallel processing architecture can easily extend to petabyte-scale

capacity for user data. Users can upgrade between sizes simply by paying the net difference in list pricing and an upgrade fee per rack. The upgrade is relatively painless; the larger system is directly loaded with data from the existing system, which is then taken down and "traded in." Typical sizes of customer deployments range from tens of terabytes to petabyte scale.

This forklift upgrade poses very low risk of service disruption and data loss, and eliminates complex data redistribution schemes, which is evidenced by the downtime which can range from seconds to a few minutes.

Global presence, reach, and services – PureData System for Analytics is marketed, sold, implemented, and supported as part of an expansive global IBM network. In particular, IBM has remained more aggressive in delivering a range of professional services around its analytic solutions, and its Global Services and Business Analytics Optimization (BAO) organizations continue to develop delivery methodology for data warehousing, BI, and advanced analytic practices. With such technology, global reach, and services resources at its disposal, IBM will figure highly in more competitive situations and continue to grab additional mind- and market share both within its existing base and with new customers – particularly at the high end of the market where the IBM brand carries considerable trust among large multinationals looking to implement regional or enterprise-wide analytics.

Weaknesses

An expensive "Rolls-Royce" solution – Buying into IBM analytics is not an inconsequential investment and the company has a tendency to invest in the features that large, complex enterprises care about. Given IBM's overall market share, most large enterprises are its committed, existing customers in some sense. While those bases do constitute a large part of the market, medium-sized organizations, with their more modest analytic needs and shallower pockets and in-house IT resources, are potentially excluded.

In-memory strategy – IBM is a relative laggard when it comes to in-memory analytics and has made no mention of this as part of its PureData announcements to date. Ovum finds it odd that a company of IBM's stature and resources has not yet articulated a clear product strategy in this area, especially considering the aggressive development of rival in-memory database offerings. IBM does have a number of products that use in-memory computing – notably its solidDB database technology. But its only in-memory solution for analytics is the legacy Informix Warehouse addition and IBM DB2 with BLU Acceleration. Most of IBM's closest analytics rivals have delivered in-memory databases, and IBM must act quickly and catch up to stay competitive.

Different configuration needed for operational analytics – IBM has consciously split out its PureData System platform into three separately licensed appliances models that are configured for specific workloads. Notably, operational analytics is supported by a different appliance – PureData System for Operational Analytics, which will especially appeal to DB2 customers. Therefore, customers requiring support for optimized transactional (OLTP) analytics workloads, typically characterized by a high percentage of interactive, ad hoc queries running directly against OLTP data, will need to purchase this appliance separately. Integration between them could prove tricky as both of these systems have vastly different software, code bases, and architecture models – PureData System for Analytics is a refresh of the Intel x86-based Netezza product line, while IBM PureData System for Operational

Analytics updates the former POWER-based Smart Analytics System 7700 – which potentially limits the portability of analytics between them, even though IBM offers a consistent administration and management layer across both. Other vendor platforms have bridged this better, allowing specific analytic use cases to be routed to the most appropriate target.

Consistency of configurations – While PureData System for Analytics comes with factory-built deployment "patterns" of the software stack that provide a degree of deployment automation, it does not necessarily guarantee a specific pattern for every conceivable analytic scenario. For broad-scale analytics, IBM's Pure Systems lines are quite heterogeneous, sharing only common self-management capabilities. That is in contrast to other analytic database vendors that utilize common components that are configured differently: for example, some have higher-capacity disks than others, and different balance of system configurations for different use cases, and remain the same regardless of whether the target is a general-purpose server, engineered system, or cloud-based. Of course, such heterogeneity (and configuration complexity) demands some degree of field support and services. IBM can provide this, but at a premium, additional cost.

Limited cloud deployment – While IBM claims that the PureData System for Analytics is well suited for cloud computing deployments, customers will find limited options for migrations to on-premise and vice versa. Deployment is currently possible in private and public clouds, but support for PaaS and IaaS is still planned for the future. That said, the system can connect to other databases that reside in the cloud, either to serve as a cache, for replication, or as a standby – via Netezza's Replication Services or using IBM's InfoSphere Information Server CDC technology. IBM also relies on partners to set up PureData System for Analytics as a cloud service.

Opportunities

Sell deeply into existing base – IBM often coexists with competitors' analytic technologies, many of whom use IBM as their data warehouse DBMS standard and become almost exclusive users of IBM solutions — these customers typically reach a point at which they exclude other top competitors from consideration due to either procurement convenience or painless and highly negotiated upgrades. This gives IBM an entrenched customer base that seems ready to use even more IBM products, which is ideal for IBM as it looks to preserve and expand revenue, margins, and share-of-wallet at those large existing customers. For example, existing IBM customers will undoubtedly be able to benefit from some kind of volume purchase agreement and will invariably have invested in some in-house expertise to effectively use and administer the system.

Positioning as part of a broad-brush analytics ecosystem – PureData System for Analytics customers will benefit from being part of a larger analytics and data management platform developed by IBM.

IBM provides a number of complementary InfoSphere-branded technology components – stream processing, data warehousing, analytic accelerators, BI tools, analytic applications, Hadoop connectivity, and data management infrastructure – that can be integrated with the system. Additionally, IBM continues to forge significant integrations with the InfoSphere stack.

Verticalization of analytics – IBM's vertical focus – backed up by a substantial professional services organization – means it can readily craft function- and vertical-specific solutions, either packaged or as

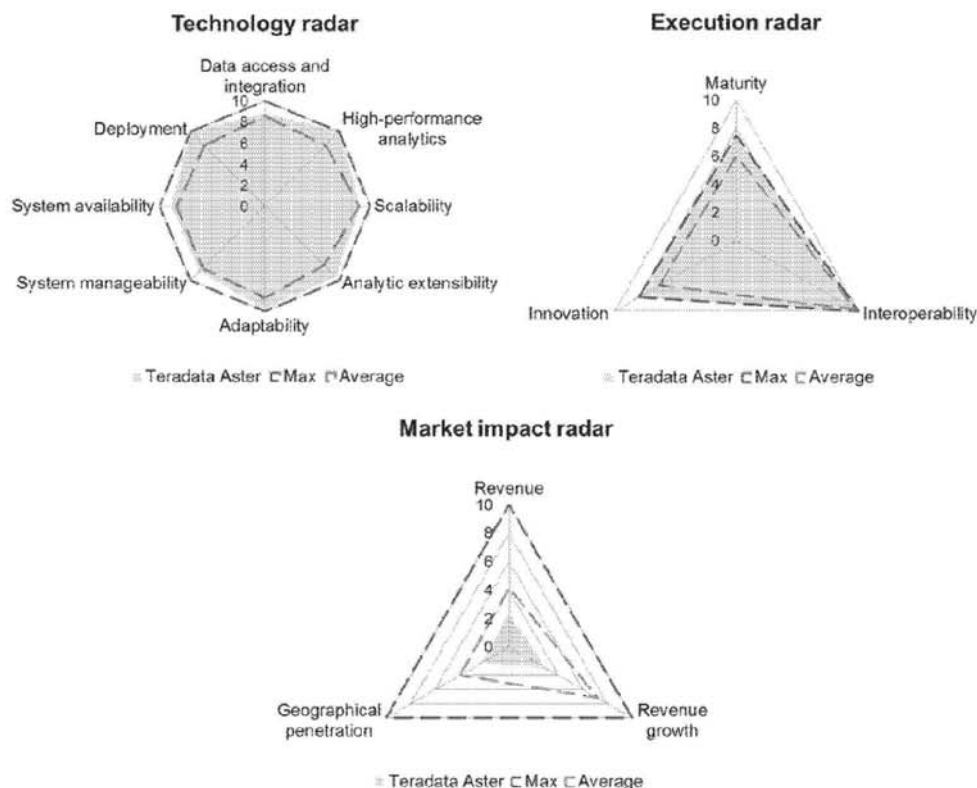
a consulting engagement. Customers can leverage the InfoSphere Warehouse "Packs" for creating customer, marketing, and campaign analytic applications. Additionally, IBM has also created mature data models for telecoms, banking, insurance, healthcare, and retail. This is also an area where IBM can leverage its unrivaled professional services divisions for customized analytic application development on top of the PureData for Analytics platform.

Threats

Limited mindshare outside of the IBM customer base – Being absorbed into a deeper and broader PureData stack reduces the visibility of the strong Netezza brand. While IBM has grown customers and channels, PureData System for Analytics win rates have fluctuated, though they have recently improved with the introduction of the N2001 version. Since IBM's partners and customer base for PureData Systems are in general relatively smaller, niche software and service providers, a key question is how effective can IBM be at selling PureData System for Analytics outside of its existing customer base. One challenge is making sure that the market knows about PureData System for Analytics and clarifying its deployment and pricing options for organizations of all sizes.

Teradata Aster Discovery Platform (Ovum recommendation: Leader)

Figure 13: Teradata Aster Discovery Platform radar diagrams



Source: Ovum

Ovum Decision Matrix: Selecting an Analytic Database, 2013–14 (IT014-002786) 14 Jan 2014

© Ovum. Unauthorized reproduction prohibited

Page 44

Ovum SWOT assessment

Strengths

Parallel SQL-MapReduce programming framework – One of the most notable features of Aster Discovery Platform is SQL-MapReduce (SQL-MR), a patented parallel programming framework that combines the analytic power of MapReduce with the familiarity of SQL. It allows SQL-savvy users to easily access, join, and analyze multi-structured sourced data and incorporate MapReduce functions into their analytic applications through a standard iterative SQL statement, without having to learn MapReduce programming or parallel programming concepts. The framework automatically parallelizes the application processing using MapReduce so that any in-database application runs in an MPP environment.

A central part of Teradata's strategy is that Aster and Hadoop nodes can work together via SQL-H for ad hoc querying, without intermediary ETL processing. With SQL-H, analysts can access archival data on Hadoop storage resources directly, easily join it with data as needed in Aster Database, and then apply the analytical power of SQL-MR. For SQL-H, Teradata takes advantage of the metadata HCatalog to improve access to data in HDFS (Hadoop Distributed File System) – a clear indication of the degree to which the company is counting on SQL-H to bring faster and easier data analysis options to Hadoop. However, SQL-H still involves two separate platforms.

Rich library of pre-built in-database analytic functions – The integrated Aster Discovery Portfolio provides more than 80 ready-to-use SQL-MR functions that are organized in modular libraries for discrete steps of the discovery process: data acquisition, data preparation, analytics, and visualization. A broad range of analytical techniques are supported including SQL, MapReduce, pattern, cluster, path, segmentation, market basket, statistical, sentiment, text, and graph (in beta right now and planned for the next release), as well as domain-specific functions such as marketing and manufacturing. These functions can be invoked and combined together in a single SQL statement and integrated into the discovery cycle. Additionally, the platform supports rich in-database modeling and analytic add-ons through the following integrations: Attensity, for text and sentiment analysis of social media content; SAS and in-Database R, for predictive modeling and scoring, which take advantage of Aster's MPP platform to score statistical models at scale; and in-database PMML execution via Zementis, which allows users to develop statistical models in the tools of their choice and score models at scale by seamlessly exporting and importing the models in the Aster Database. Developers can also use the SQL-MR framework to write their own expressive analytic functions in a variety of languages, including Java, C, C++, C#, Python, and R, and push them into the Aster Database for advanced in-database analytics. In the 5.10 release, Teradata has also added a new category of visual SQL-MR functions for advanced visualizations, including flow, affinity, and hierarchy visualizers. In the future, Aster will have a more seamless integration with the R client in supporting in-database R modeling. Teradata is also considering support for additional R distributions besides the base open source R.

Strong data connectivity, particularly to Hadoop – Given the investment put into the SQL-MapReduce framework, it is not surprising that Teradata boasts one of the industry's deepest stack integrations of the Aster Database into the Apache Hadoop environment – the Aster Database comes with an integrated development environment for developing and testing MapReduce functions

and SQL statements. It also includes a Data Acquisition Module that not only delivers out-of-the box adapters and connectors to access the Teradata Database and other relational databases, semi-structured data (e.g. Web logs, XML documents), unstructured data (e.g. call-center IVR records, BLOBs, geocoded data), ODBC and JDBC certifications for BI, and visualization and ETL tools, but also deep connectors for Apache Hadoop and Teradata's own SQL-H – which enables standard SQL and SQL-MapReduce queries to be run against data stored in Aster and Hadoop by leveraging Apache HCatalog. This means that business analysts can directly analyze vast amounts of Hadoop data without needing programming skills or an understanding of how data is stored within the HDFS.

Engineered for high-scale and fast query performance – Aster Discovery Platform is based on a highly scalable MPP architecture that parallelizes the execution of data loads, exports, and queries to take full advantage of all available hardware resources. A 40Gbps InfiniBand network connects the Aster and Hadoop nodes. This fast, dedicated, and reliable network interconnect can also be used to connect multiple Aster systems.

The embedded SQL-MapReduce framework uniquely enables optimized analytic processing of large data volumes by automatically parallelizing processing of iterative query workloads and analytic applications – with the MapReduce functions, the execution of complex schema-on-database-read logic can be parallelized. To safeguard query performance, the solution does query optimization and execution across SQL and SQL-MR execution in one single query. The optimizer in Aster Database is mature and can handle multi-table joins efficiently, including queries that access Hadoop data. It provides a number of features, including global and local optimizers for both SQL and SQL-MapReduce analytic queries; hash distribution of data; range-, list-, and value-based partitioning of data; and indexing support for fast lookups. Teradata is also beefing up dynamic workload capabilities for better resource allocation and contention at higher concurrencies as well as columnar data layouts for faster access to highly projective queries and analytics.

Part of a unified architecture – Teradata Aster Discovery Platform forms an integral part of what the company has dubbed its Unified Data Architecture (UDA), which represents Teradata's vision for the logical data warehouse. While the aim of the UDA is to align the best technology to specific analytic needs, architecturally, it is about ensuring that all of Teradata's analytic database platforms work in harmony with one another, including the Teradata Database-based integrated EDWs and data marts, Teradata Aster, and the Apache Hadoop data-processing framework as a data-staging and ingestion point. The benefit for customers is twofold: they get to utilize common components that are configured differently (in terms of capacity and analytic use cases); and can tap into core Teradata database-enabling technologies such as the Teradata Bynet backplane, Viewpoint, and Unity products for analytic portability across different and appropriate target platforms, and unified system management and node-monitoring benefits. Aster Discovery's support for Apache HCatalog is also starting to address metadata management, which is critical for logical data warehousing architectures and helps Teradata make it easier to share and reuse data stored in external file systems, such as HDFS.

Market presence and loyal customer base – The Teradata brand is highly trusted and well vested among its 1,200 customer base. Ovum finds customers to be extremely loyal to the Teradata platform – partly due to the high investment they have made – and have used and upgraded to various products for periods up to a decade, attracted to the expanded architectures of Teradata's appliances. Teradata

also demonstrates a durable track record in customer satisfaction and (local) support across a diverse range of industries and geographies.

Weaknesses

Relies on having a strong data movement infrastructure in place – SQL-H's strength is that it takes Hadoop data into an SQL environment (where the SQL is superior and performance is superior to Hadoop). But it is also its weakness, since it relies heavily on data movement to analyze Hadoop data. However, organizations do have the flexibility to persist and store the Hadoop data in the Aster Database. The vendor believes that analytics done inside Hadoop alone are not optimal for many use cases. Instead, Teradata Aster Database combines and analyzes diverse data from transactions, interactions, and sensors to do rapid and interactive exploration. The platform can access data from such different sources and perform rapid iterations using its SQL-MapReduce framework and massively scalable architecture. The vendor also provides an extensive portfolio of pre-built functions that automate these tasks for business analysts. Overall, this is not a marked weakness but a conscious strategy to use a familiar SQL environment to do NoSQL tasks, which has its own advantages and disadvantages. Most importantly, Ovum believes that this process of analyzing data will often require higher-speed interconnects to improve performance by brute force (to achieve meaningful load speeds), which may make it expensive in terms of overheads compared to competing vendors' offerings that analyze all data inside Hadoop.

Varying deployment cycles – Teradata systems are certainly not implemented cheaply as they are typically MPP integrated data warehouse systems performing analytics at scale to meet the needs of organizations, and Ovum has some concerns over total cost of ownership. The Aster Discovery Platform is, however, designed as a purpose-built appliance solution for Big Data analytics and discovery. While deployments for large number of users is not typical – usually up to 100 users – even in very large enterprises, it requires some dedicated level of in-house IT staffing and expertise to run and maintain. For example, a dedicated server is often needed (and recommended) for special-purpose managed server nodes. Deployment times vary, depending on various factors relating to appropriate data access, connectivity to the customer's internal IT infrastructure, and the assumption of meaningful analytic business use cases. If these are not in place, deployment times can increase. Aster Discovery Platform does, however, provide the Discovery Portfolio of pre-built functions to speed up implementations and allow organizations to benefit from analytics quicker.

No support for streaming and geospatial data types – At present, Aster Discovery Platform does not provide native support for fast, streaming data, and relies heavily on partners with streaming engines that can be used to stream data into the Aster Discovery Platform. Nor does it provide strong support for geospatial data types (it can only take advantage of geospatial data through custom SQL-MapReduce functions). While streaming data is typically used in transaction-processing applications, considerable benefit can also be gained in analytic environments for operational and/or realtime discovery. And with the advent of increased mobility and sensor-based data, location sensitivity is playing an increasingly important role in many types of analytics and discovery insights.

Dearth of Teradata and Aster skills on the market – Teradata risks offering advanced analytic products and platforms to a mainstream market that does not have an adequate ability to support them. Ovum still generally finds a lack of skilled Teradata professionals, relative to other vendor platforms,

which means that organizations have to pay premium rates to scale internal IT teams to support Teradata platforms. That carries the risk of customers delaying upgrades to newer versions of its platform. However, customers can turn to Teradata partners such as MicroStrategy and Tableau, which have built more integrated support for Teradata Aster and offer easier-to-use implementations that are interactive and visually designed for analysts without SQL. Teradata could also do more to ensure that third-party DBA tools more easily work with Teradata platforms.

Opportunities

Market shift away from warehouse-driven analysis to ad hoc discovery platforms – Teradata has over 30 years' experience of solving BI and analytic problems with its EDWs. With increasing interest in exploratory analytic platforms that complement these EDWs, Ovum sees an opportunity for Teradata to win new customers from this base of newcomers as they develop additional requirements. Significantly, Teradata has focused on providing pre-built two-way connectors for direct, parallel data extraction or querying from the Aster Discovery Platform to the Teradata EDW.

Provide thoroughbred horses for the right courses – Teradata Aster takes a different approach to designing its hybrid MPP columnar SQL/Hadoop appliance. Instead of offering a choice of modules, Teradata Aster provides a standardized appliance, but with different configurations internally for nodes handling the varying workloads.

Threats

Explaining and ironing out potential overlaps – Teradata seems to be using its Aster platform to place a stake in the "middle ground" between its core Teradata Database and Hadoop environments. But as Hadoop performance improves in the long run, that could possibly be a long-term threat to Aster, leading to the question of whether Aster's capabilities might ultimately be subsumed into the core Teradata Database. Teradata therefore continues to face the challenge of explaining the potential overlap between the Teradata Database, Aster technology, and Hadoop, although it has gotten better at articulating its positioning of late. Likewise, the company has generally struggled to find the right term, and positioning, for the Aster Discovery Platform. Part of that is the positioning of the Aster Database alongside Hadoop; Teradata sees the two as complementary, and supported by relationships with Hortonworks and Cloudera. However, many customers looking to do interactive SQL analysis on data stored in Hadoop will also be looking at Cloudera's Impala project and the MapR-initiated Apache Drill.

Competition on both fronts – Teradata's primary competition comes from the database industry heavyweights – IBM, Oracle, Microsoft and, more recently, EMC and SAP – that are positioned as more general-purpose analytic platforms. Aster Discovery Platform might therefore be pigeonholed alongside specialized analytic platforms that Aster competed with prior to its acquisition by Teradata – such as ParAccel and HP Vertica, both of which are also positioned for similar exploratory workloads. Additionally, customers looking for hybrid SQL and MapReduce platforms also have other options in hand: EMC Pivotal's Modular Data Computing Appliance (which combines EMC's Greenplum Database and Greenplum HD Hadoop distribution in a single appliance), Hadapt's Adaptive Analytical Platform, and RainStor's Big Data Analytics on Hadoop.

Remains a potential takeover target – As the largest analytic database specialist, Teradata is the number one target for a variety of big-name data management providers, not least Oracle, IBM, and EMC, especially as market attention continues to increase on trying to build SQL engines on Hadoop.

APPENDIX

Further reading

Analytic Databases: Technology to Solve Modern Decision-Making Challenges, IT014-002699 (March 2013)

Choosing SQL Analytic Database Appliances, Part 2, IT014-002787 (September 2013)

Choosing SQL Analytic Database Appliances, Part 1, IT014-002777 (August 2013)

Who's Who in Outsourced Analytics, 2013, IT014-002780 (August 2013)

Author

Surya Mukherjee, Senior Analyst, Software – Information Management

surya.mukherjee@ovum.com

Madan Sheina, Lead Analyst, Software – Information Management

madan.sheina@ovum.com

Ovum Consulting

We hope that this analysis will help you make informed and imaginative business decisions. If you have further requirements, Ovum's consulting team may be able to help you. For more information about Ovum's consulting capabilities, please contact us directly at consulting@ovum.com.

Disclaimer

All Rights Reserved.

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior permission of the publisher, Ovum (an Informa business).

The facts of this report are believed to be correct at the time of publication but cannot be guaranteed. Please note that the findings, conclusions, and recommendations that Ovum delivers will be based on information gathered in good faith from both primary and secondary sources, whose accuracy we are not always in a position to guarantee. As such Ovum can accept no liability whatever for actions taken based on any information that may subsequently prove to be incorrect.